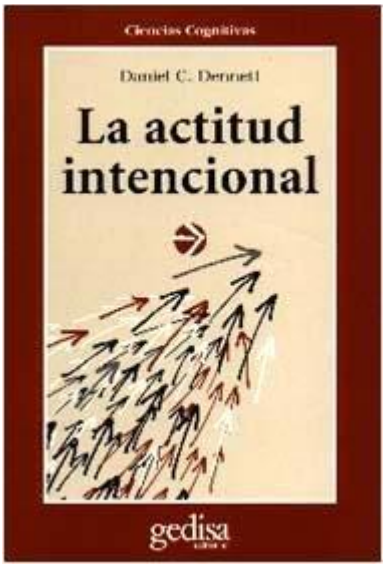
Daniel C. Dennett

LA ACTITUD INTENCIONAL



Daniel C. Dennett

LA ACTITUD INTENCIONAL

Serie CLA • DE • MA
CIENCIAS COGNITIVAS

LA ACTITUD INTENCIONAL

por

Daniel C. Dennett

gedisa
editorial

Título del original en inglés: *The Intentional Stance*

Traducción: Daniel Zadunaisky

Diseño de cubierta: Marc Valls

Segunda edición, marzo de 1998, Barcelona

Muntaner, 460, entlo., 1.ª
Tel. 201 60 00
08006 - Barcelona, España
*e-mail:*gedisa@gedisa.com
http://www.gedisa.com

ISBN: 84-7432-395-9
Depósito legal: B-9.797/1998

Impreso en Limpergraf
c/ del Río, 17 - Ripollet

Impreso en España
Printed in Spain

Dedicado a la memoria
de Basil Turner,
vecino, amigo y maestro.

Indice

PREFACIO 11

1. Arrancando con el pie derecho 15
El sentido común y el punto de vista de la tercera persona (16); La ciencia popular y la imagen manifiesta (20).

2. Los verdaderos creyentes: La estrategia intencional y por qué funciona 25
La estrategia intencional y cómo funciona (27); Los verdaderos creyentes como Sistemas Intencionales (33); ¿Por qué funciona la estrategia intencional? (42).

Reflexiones: Modelos reales, hechos más profundos y preguntas vacías 44

3. Tres clases de psicología intencional 50
La psicología popular como fuente de teorías (50); La teoría del sistema intencional como una teoría de competencia (62); La psicología cognitiva subpersonal como teoría de ejecución (65); Las perspectivas de la reducción (69).

Reflexiones: El instrumentalismo reconsiderado 72
El instrumentalismo (74).

4. Comprendiéndonos a nosotros mismos 83

Reflexiones: Cuando las ranas (y otros) cometen errores 98
El error del vendedor de limonada (98); Psicología de la rana (101); Las ilusiones del realismo (105)

5. Más allá de la creencia 111
Las actitudes proposicionales (114); Actitudes oracionales (122); Actitudes nocionales (140); El *de re* y el *de dicto* desmantelados (160).

Reflexiones: Acerca de la acerquidad 184
Las proposiciones (185); Los mundos nocionales (189); El principio de Russell (190); El *de re / de dicto* (190).

6. Los estilos de representación mental 192
Reflexiones: El lenguaje del pensamiento reconsiderado 203

7. Los sistemas intencionales en la etología cognitiva: Defensa del "Paradigma panglossiano" 211
La teoría del sistema intencional (214); Cómo usar la evidencia anecdótica: el método de Sherlock Holmes (222); Una perspectiva biológica más amplia de la actitud intencional (228); Defensa del "Paradigma planglossiano" (231).
Reflexiones: Interpretando a los monos, los teóricos y los genes 238
Los ancestros y la progenie (238); Reconsideración del paradigma panglossiano (245); El adaptacionismo como interpretación radical retrospectiva (250).

8. La evolución, el error y la intencionalidad 254
El caso del ordenador errante de dos bits (256); El diseño de un robot (261); Leyendo la mente de la Madre Naturaleza (264); El error, la disyunción y la interpretación inflada (266); ¿Está la función en la mirada del observador? (277).

9. El pensamiento veloz 286

10. Examen de mitad de curso: Comparación y contraste 299

BIBLIOGRAFÍA 309

INDICE TEMÁTICO 327

Prefacio

La teoría de la intencionalidad que se presenta en este libro ha venido evolucionando progresivamente desde hace más de veinte años. Si bien las ideas principales fueron expresadas de manera rudimentaria en *Content and Consciousness* en 1969, fue la publicación en 1971, de *Intentional Systems* la que inició la serie de artículos acerca de lo que yo llamo la actitud intencional y los objetivos que se descubren a partir de esa actitud: los sistemas intencionales. Los primeros tres de estos artículos (Dennett, 1971, 1973, 1976b) se reimprimieron en *Brainstorms* en 1978, y tanto los críticos como los estudiantes tratan este libro como la expresión canónica, como la meta de mi teoría. Sin embargo, descubrí enseguida que la defensa de mi posición crecía como reacción ante la crítica, y así fue como me sentí obligado a escribir una serie de ensayos pos*Brainstorms* en los que intenté corregir, volver a expresar y ampliar mi punto de vista.

Sin embargo, la mayor parte de estos ensayos se dispersaron en volúmenes relativamente inaccesibles, gracias al efecto inexorable de la "Gravedad del reflector": a medida que las ideas de uno se convierten en "Centro de Interés", se le invita a tomar parte en cada vez más conferencias, que absorben todo lo que uno ha publicado para editarlo con atraso en actas de conferencias y antologías de interés especializado. No queda nada que presentar en revistas de opinión para su lectura inmediata. El objetivo de este libro es superar los efectos secundarios negativos de esa difusión tan gratificante en otro sentido.

Seis de esos ensayos dispersos se reimprimieron en este volumen (capítulos 2 al 7), encabezados por un ensayo acerca de sus aspiraciones y presunciones metodológicas, unidos por reflexiones y seguidos por dos ensayos nuevos (capítulos 8 y 9), en los que los temas y argumentos de los capítulos precedentes convergen en reclamos bastante sorprendentes acerca de la relación entre la evolución, el diseño cerebral y la intencionalidad. El capítulo 10 es el intento de adoptar la actitud que tendría un observador imparcial sobre mi propio trabajo y describir el lugar que ocupa en la evolución del pensamiento actual acerca de la "intencionalidad de los estados mentales".

Este libro no presenta la totalidad de mi teoría acerca de la mente, sino sólo, podríamos decir, la primera mitad: el contenido. La otra mitad, la conciencia, necesita también de un segundo relato (la parte tres de *Brainstorms* fue el primero), pero para eso hará falta otro volumen, al que me estoy dedicando actualmente. A la conciencia se la considera habitualmente, en es-

pecial por aquella gente que está fuera del campo de la filosofía, como el desafío más notable (y más desconcertante) a las teorías materialistas acerca de la mente. No obstante, es muy raro que la mayoría de las personas más importantes que participan en los debates *acerca del contenido mental*, a quienes este volumen está especialmente dirigido, hayan mantenido un silencio conspicuo acerca del tema de la conciencia. No se encuentra ninguna teoría, o siquiera el esbozo de una teoría acerca de la conciencia en los escritos de Fodor, Putnam, Davidson, Stich, Harman, Dretske o Burge, por ejemplo. Por otra parte yo sí tengo una teoría acerca de la conciencia (y siempre me costó entender cómo los demás suponen que pueden ignorar o postergar el tema), pero la última versión es demasiado burda como para ser incluida en este volumen. Aquel que esté impaciente por conocer la nueva versión de esta segunda mitad de mi teoría de la mente puede acercarse a ella a través de las ideas expresadas en los ensayos ya publicados y de los próximos: *How to Study Human Consciousness Empirically; Nothing Comes to Mind* (1982b), *Why Do We Think What We Do About Why We Think What We Do?* (1982d), *Reflection, Language and Consciousness (Elbow Room*, 1984d, págs. 34-43), *Julian Jaynes' Software Archeology* (1986d), *Quining Qualia* (próximo a aparecer d) y *The Self as the Center of Narrative Gravity* (por aparecer g).

Otro desafío, también comúnmente considerado insuperable por las teorías materialistas acerca de la mente, es el problema del libre albedrío. Le he dedicado un libro aparte a ese desafío, *Elbow Room*, de manera que el tema rara vez volverá a tocarse en estas páginas. Si hay otras objeciones importantes a mi teoría, todavía no he tenido conocimiento de ellas.

Puesto que los ensayos de este volumen, que ya fueron publicados, aparecieron en el transcurso de cinco años signados por su parte de controversia, incomprensión y corrección, no es sorprendente que hayan sido pocos los que pudieron discernir la posición equilibrada resultante. A veces se ha hablado de mí, como alguien que les presenta a sus críticos un blanco móvil. Algo de cierto hay en eso. Me dispuse a aprender de mis errores y a retractarme de exigencias imprudentes. Sin embargo, el movimiento es relativo, y, cuando a un observador algo le parece proteico y errático, puede que sea porque el observador apenas ha comenzado a adivinar la forma que ese algo siempre tuvo. Hace poco tiempo un neurocientífico me felicitó por revisar una teoría que había defendido desde *Content and Consciousness* en 1969, una experiencia meditada y tranquilizante que me llevó a revalorar mi estrategia expositora con el paso de los años. Cuando releo ese libro, ahora publicado en rústica después de haber estado agotado durante varios años, me impresiona más mi constancia doctrinaria que mi evolución. La mayor parte de los cambios me parecen ampliaciones, extrapolaciones, argumentos adicionales y no cambios. Sea como fuere, es probable que yo haya subestimado en exceso la posibilidad para tomar el camino equivocado que hay en mi estilo juguetón y poco sistemático. Por tanto, en este libro hago todo lo posible por hacer una pausa, poner mis carretas en círculo y presentar y defender mi opinión de manera más ordenada.

Algunas de las partes previamente inéditas de este libro están tomadas de mis clases en la cátedra Gavin David Young de la Universidad de Ade-

laide en 1984, de las pronunciadas en la Ecole Normale Supérieure de París en 1985, y de otras disertaciones realizadas en distintas reuniones, conferencias y coloquios en los últimos dos o tres años, que me hicieron adquirir la comprensión que se ha concretado en este volumen.

Tengo mucha gente a quien agradecer sus consejos y sus críticas acerca de los primeros borradores del material inédito que aparece en este volumen: especialmente a Kathleen Atkins, quien no sólo ayudó a organizar y redactar todo el libro, sino que además me convenció para que hiciera correcciones más importantes en la presentación y defensa de mis ideas; y a Peter Bieri, Bo Dahlbom, Debra Edelstein, Doug Hofstadter, Pierre Jacob, Jean Khalfa, Dan Lloyd, Ruth Millikan y Andrew Woodfield. Me complace mucho incorporar mi voz al coro ya tradicional de alabanzas dirigidas, por tantos autores, a mis buenos amigos Harry y Betty Stanton, quienes a través de los años me hicieron sentir tan orgulloso de ser uno de los autores publicados por Bradford Books. Y gracias, como siempre, a mi esposa Susan por su apoyo y paciencia, y a mis colegas de Tufts.

Universidad de Tufts
Enero de 1987

1

Arrancando con el pie derecho

Para mucha gente, hablar de la mente es como hablar de sexo: ligeramente embarazoso, indecoroso y hasta deshonroso. "Claro que existe", dirán algunos, "pero ¿es necesario que hablemos de ella?". Sí, lo es. Muchos preferirían hablar del cerebro (que, después de todo *es* la mente) y querrían creer que todas las cosas maravillosas que tenemos que decir acerca de la gente se podrían decir sin caer en una charla *mentalista*, vulgar e indisciplinada, pero en este momento está muy claro que muchas cosas que deben decirse, no pueden ser dichas en los lenguajes restringidos de la neuroanatomía, la neurofisiología o la psicología conductista. No son sólo las artes y las humanidades las que tienen que hablar de la mente; las diversas tentativas puritanas de dar por terminadas las ciencias biológicas y sociales sin siquiera referirse a ella, ya han revelado ampliamente su futilidad.

En realidad, hay cierta aproximación para un nuevo consenso entre los científicos cognitivos y los neurocientíficos más liberados sobre el hecho de que puede haber —de algún modo debe haber—, una ciencia materialista responsable no sólo del cerebro, sino también de la mente. Sin embargo, todavía no hay consenso acerca de cómo se ha de manejar esta ciencia responsable de la mente.

Este libro trata de cómo hablar de la mente. Es un libro filosófico, escrito por un filósofo, y que se ocupa en especial de los temas a medida que han ido apareciendo en la literatura filosófica, pero no está dirigido sólo a los filósofos. Quienes se dedican a otras disciplinas pero están siempre ansiosos o por lo menos dispuestos, aunque no sea de muy buena gana, a gratificarse con distintos tipos mentalistas de discusión, descubren que los filósofos, que nunca tuvieron vergüenza de hablar de la mente, tienen mucho que decirles acerca de cómo hacerlo. En realidad, nosotros, los filósofos, tenemos en realidad demasido que decir. Sólo una pequeña parte de lo que hemos dicho tendría alguna posibilidad de ser cierta o útil, e incluso hasta lo mejor puede ser mal interpretado. La filosofía no produce a menudo "resultados firmes y dignos de confianza como lo hace la ciencia en sus mejores momentos. Puede, no obstante, producir nuevas maneras de ver las cosas, de pensar en ellas, de formular las preguntas, y de ver qué es lo importante y por qué".

Puesto que todos los que se están ocupando de la mente se ven acosados por problemas tácticos acerca de qué preguntas tratar de contestar, ésta puede ser una contribución valiosa. Todos nos enfrentamos con fenómenos desconcertantes; ¿qué podría ser más conocido, y al mismo tiempo más misterioso que una mente? Tenemos también una cantidad arrolladora de datos

acerca del objeto más complejo que se ha encontrado en el universo: el cerebro humano, y acerca de la enorme variedad de conductas que el cerebro es capaz de modular. Finalmente, nos deja perplejos una multitud entrometida de intuiciones persistentes que vienen Dios sabe de dónde. Por tanto, los teóricos de todos los campos corren el riesgo de seguir a sus distinguidos predecesores y arrancar con el pie equivocado por culpa de algún error de concepto filosófico acerca de la naturaleza de los fenómenos, la gama disponible de opciones teóricas, la configuración de los trabajos teóricos, o los requisitos que hay que cumplir para dar una explicación afortunada de la mente.

No hay manera de evitar tener preconceptos filosóficos; la única opción es estudiarlos o no en forma explícita y cuidadosa en algún momento de nuestra tentativa. Es posible, por supuesto, que algunos de los teóricos actuales que carecen de cultura filosófica tengan la suerte suficiente como para atesorar sólo los preconceptos filosóficos más acertados; quizá la atmósfera de la época asegure esto aun sin tener comunicación directa con los filósofos. Y por cierto: hay que tener presente que algunos de los más perniciosos preconceptos del pasado han sido poderosos legados de la filosofía académica que los científicos han interpretado mal, a causa de un entusiasmo y simplificación exagerados. Nos viene a la memoria el positivismo lógico y, más recientemente, la inconmensurabilidad de los paradigmas kuhnianos. No obstante, los filósofos creemos que podemos ayudar, y nos gratifica encontrar un número cada vez mayor de gente que acude a nosotros en busca de ayuda con una actitud, ciertamente adecuada, de escepticismo cauteloso.

Este libro presenta las bases de mi teoría de la mente: mi explicación de la actitud intencional. Quienes están familiarizados con esa explicación encontrarán pocas innovaciones importantes de la teoría, pero sí algunas innovaciones en su exposición y defensa, en especial en los comentarios que siguen a cada uno de los ensayos reimpresos, donde intento aclarar y ampliar mis argumentos previos. El último capítulo está dedicado a una comparación sistemática de mi punto de vista con aquellos que fueron defendidos hace poco tiempo, utilizando las críticas y objeciones de otros para encarar los puntos problemáticos. En estos nuevos ensayos he tratado de presentar y responder a todas las objeciones que a mi explicación se publicaron; de corregir los malos entendidos y las malas interpretaciones. También explico, de paso, algunos de los puntos principales de coincidencia y desacuerdo con otros autores que han escrito sobre estos temas, y señalo algunas implicaciones, en general no admitidas, de mi posición sobre controversias actuales.

La presentación básica de mi teoría de la actitud intencional se encuentra en el próximo capítulo, *True Believers:* los verdaderos creyentes, con el que ahora intento reemplazar a *Intentional Systems* como la expresión más avanzada de mi opinión. En el resto de este capítulo retrocedo algunos pasos y comento algunas hipótesis no discutidas de los otros ensayos.

El sentido común y el punto de vista de la tercera persona

Aquí, en el planeta Tierra, hay formas de vida muy complicadas. El sentido común nos dice que muchas de ellas tienen vidas mentales —men-

tes— de tipos confusos de prever. Lo que el sentido común nos dice no es suficiente. No sólo deja sin resolver demasiados problemas apremiantes, sino que se entrega con frecuencia a intuiciones persuasivas que se contradicen. Desde algunas posiciones ventajosas es "evidente" que los animales de sangre caliente tienen mentes como las nuestras, mientras que los insectos parecen ser "meros autómatas". Desde otros lugares de privilegio la diferencia entre nosotros y el chimpancé parece mayor que la que hay entre una paloma y un robot. La idea de que ningún autómata podría ser consciente tal como nosotros lo somos está totalmente popularizada, pero se la puede hacer parecer sospechosamente parroquial y carente de imaginación, un ejemplo de ilusiones descaminadas. Algunas patologías aparentemente probadas de la mente y el cerebro humanos son tan contraintuitivas que detallarlas provoca a menudo que sean descartadas con sorna. Hace poco tiempo, una de mis alumnas le transmitió a su profesor de literatura el relato que yo había hecho en clase de las patologías raras pero bien estudiadas de la negación y hemidescuido de la ceguera. el profesor le aseguró con firmeza que yo lo había inventado todo, que debía haber estado haciendo algún experimento sobre la credulidad de mis alumnos. Para él estaba claro que el profesor Dennett estaba inventando otra de sus extravagantes fantasías de ciencia ficción, un sondeo intuitivo más para embaucar a los crédulos. Cuando tantos "hechos evidentes" compiten entre sí, el sentido común no es suficiente.

No hay reglas que rijan la manera en que nosotros, los teóricos, debemos apelar al sentido común. De uno u otro modo debemos partir de la base del sentido común si esperamos ser comprendidos o comprendernos. Pero la confianza en cuaquier ítem especial de sentido común es traicionera: lo que para una persona es un fundamento sólido para otra es un vestigio falsamente convincente del punto de vista de un mundo perimido. Aun si algunos aspectos de lo que pasa por ser el sentido común son la Verdad resplandeciente e inmutable, es probable que otros no sean más que las ilusiones cognitivas de nuestra especie, abrumadoramente persuasivas para nosotros debido a la existencia de ciertos atajos en el diseño de nuestros sistemas cognitivos. (A una polilla fototrófica le puede parecer una verdad *apriorística* que siempre está Bien encaminarse hacia la Luz; no concibe ninguna otra alternativa.) Otras formas de expresión del sentido común no son más que versiones popularizadas y atenuadas de la ciencia de antaño.

Clasificar estos aspectos del sentido común en verdaderos, falsos, engañosos e indignos de confianza es un buen trabajo para un filósofo. En realidad los filósofos se especializan en este tipo de tareas. Algo que hemos aprendido de los distinguidos fracasos del pasado es que ésta no es una tarea sistemática, accesible a un enfoque puramente básico o axiomático. Más bien tenemos que arremeter cuando sea oportuno e intentar alcanzar una visión estable oponiendo entre ellas una gama de intuiciones, descubrimientos y teorías empíricas, argumentos rigurosos y experimentos imaginativos del pensamiento.

Algunas escaramuzas útiles de esta campaña consisten ciertamente en exploraciones rigurosas y formales de conjuntos especiales de presentimientos. Esa es en realidad la mejor luz para observar los distintos fracasos formalistas de la filosofía, como si estuvieran prologados por la siguiente pregun-

ta: "¿Qué tal si formuláramos *estas* hipótesis y avanzáramos según *estas* restricciones?". Como dice Fodor: "Muy a menudo la forma de una teoría filosófica es: *Probemos buscar por acá*" (1981a, pág. 31). En filosofía todos los sistemas formales deben estar "motivados", y la tarea informal de proporcionar dicha motivación contribuye más a la claridad filosófica (o por lo menos a una doctrina) que el sistema al que le allana el camino. Siempre hay más de un sistema que sea candidato o una perspectiva que claman por ser objeto de exploración y evolución filosóficas, y en un campo del pensamiento tan indisciplinado, las consideraciones tácticas desempeñan un papel desusadamente importante. Estas consideraciones tácticas se disfrazan a menudo, sin embargo, de principios primordiales.

Por tanto comienzo con una elección táctica. Declaro que mi punto de partida es el mundo objetivo, materialista, tal como lo ve la tercera persona de las ciencias físicas. Esta es la elección ortodoxa de la actualidad del mundo filosófico angloparlante, pero que tiene sus detractores, el más notable de los cuales es Nagel, que ha dedicado un libro, *The View from Nowhere* (1986), a deplorar los efectos de esta elección táctica. Puesto que el punto de partida de Nagel es una alternativa más importante, en comparación con la mía, comparémoslas brevemente por si estuviéramos pasando algo por alto.

No estoy seguro de que Nagel sea uno de los que creen que pueden *probar* que mi elección es un error, pero es cierto que él *afirma* que lo es.

> Hay cosas acerca del mundo, la vida y nosotros mismos que no se pueden entender bien desde un punto de vista de objetividad máxima, por más que dicha objetividad logre extender nuestra comprensión más allá del punto del que partimos. Una gran parte está conectada en forma esencial a determinado punto de vista, o tipo de punto de vista, y la tentativa de dar una explicación completa del mundo en términos objetivos, separados de estas perspectivas, lleva, de manera inevitable, a reducciones falsas o a una negación total de que ciertos fenómenos evidentemente reales existen (pág. 7).

Mis intuiciones acerca de lo que "no se puede entender bien" y lo que es "evidentemente real" no coinciden con las de Nagel. Nuestros gustos son muy diferentes. Por ejemplo, a Nagel le abruma el deseo de desarrollar una explicación evolutiva del intelecto humano (págs. 78-82). A mí, esa perspectiva me llena de gozo. Mi sentido de que la filosofía está aliada a, y es sin duda una continuación de las ciencias físicas sustenta tanto mi humildad acerca del método filosófico como mi optimismo acerca de su progreso. Para Nagel, esto es puro cientificismo.

> Hasta el punto en que esas teorías importantes tengan validez, ellas simplemente amenazan con empobrecer el panorama intelectual durante un tiempo al inhibir la expresión seria de ciertos temas. En nombre de la liberación, estos movimientos nos han traído represión intelectual (pág. 11).

Nagel es valiente e inteligente a la vez. Se necesita valor para defender el misterio, e inteligencia para ser tomado en serio. Nagel avisa una y otra vez que no tiene respuestas para los problemas que plantea, pero que prefiere su mistificación a los esfuerzos demistificadores de otros. Por extraño que pa-

rezca, Nagel estaría de acuerdo conmigo en que su punto de partida táctic no sólo produce perplejidad, sino que es una clase de perplejidad para la qu él mismo no ofrece ninguna escapatoria. Para mí, ese callejón sin salida e equivalente al *reductio ad absurdum* de su método, pero Nagel recomienda con valentía aceptar el resultado:

> Ciertas formas de perplejidad —por ejemplo, acerca de la libertad, el conocimiento y el sentido de la vida— encierran, en mi opinión, más comprensión que cualquiera de las supuestas soluciones de esos problemas (pág. 4).

Nagel es el defensor contemporáneo más elocuente de los misterios, y cualquiera que sospeche que yo he subestimado los problemas que propongo para mi teoría, se sentirá fortalecido por las afirmaciones, en contra de Nagel. Afirmaciones, no argumentos. Puesto que Nagel y yo partimos de distintas perspectivas, sus argumentos dan por sentada su oposición en contra de la mía: lo que él considera completamente claro y sin necesidad de un sustento mayor, a mí no me impresiona. Supongo que, cualquiera que resulte ser la verdadera teoría de la mente, derribará algunas de nuestras convicciones previas, por tanto no me importa que me definan las implicaciones contraintuitivas de mi punto de vista. Cualquier teoría que progrese está destinada a ser inicialmente contraintuitiva. Sin duda Nagel, que dice que su libro es "deliberadamente reaccionario", se mantiene igualmente firme cuando se le hace notar que su fidelidad a ciertas intuiciones es lo que le impide escapar de su perplejidad por distintos caminos promisorios de la investigación científica.

Por tanto el sentimiento es mutuo; damos por sentado que una opinión está en contra de la otra. No presupongo que un punto de partida alternativo como el de Nagel deba ser erróneo, y que todo lo que vale la pena tomarse en serio en el Universo deba ser accesible desde mi punto de partida. Me siento impresionado, sin embargo, por su probado rendimiento de comprehensión (aparente) y más aun por su promesa de dar frutos futuros.

Nagel alega demostrar que la tentativa de conciliar lo objetivo con lo subjetivo, es "esencialmente imposible de llevar a cabo" (pág. 4), o podría tener razón —aunque yo no esté en absoluto convencido—. Sin embargo, hay quienes le acompañan en la sospecha de que hay algo sutilmente incoherente en la visión más o menos estándar que los científicos tienen del mundo y de nuestro lugar en él —para algunos conflictos insolubles entre lo objetivo y lo subjetivo, lo concreto y lo abstracto, lo macro y lo micro— (véase Dennett 1984d, págs. 128-29). ¿Acaso el objetivista autodesignado no depende furtivamente de algún compromiso previo con puntos de vista irreductibles? ¿O el propósito de "reducir" estos puntos de vista a la biología, la química o la física no se autodestruye de todos modos? Se murmura que allí abajo en los subsótanos de la física contemporánea los alquimistas modernos están volviendo a convertir el materialismo en idealismo.[1] Las partículas de quantum *parecen* ser verdaderamente, a veces como lo ha dicho David Moser, "los sueños de que está hecha la materia".

Tal vez quienes desconfían de las suposiciones y aspiraciones francamente materialistas de la actual imagen científica tengan razón en hacerlo,

pero yo lo dudo, y opto por no enfrentarme a sus sospechas con vehemencia ya desde el principio. La ortodoxia actual de mi punto de partida científico *podría* deberse tanto a factores políticos y sociales como a cualquier justificación filosófica. Aunque yo no lo crea, puedo ver lo que hay de plausible en el diagnóstico de Nagel: "Es como el odio de la infancia y resulta en un esfuerzo vano por crecer antes de tiempo, antes de que uno haya pasado por las confusiones formativas esenciales y las esperanzas exageradas que hay que experimentar cuando se va hacia la comprensión de algo" (pág. 12).

Mi presentimiento táctico, sin embargo, es que aun cuando esto sea así, la mejor manera de llegar a entender la situación es empezar aquí y dejar que las revoluciones que estén por producirse se produzcan desde adentro. Por lo tanto, propongo ver *cómo es la mente desde la perspectiva materialista*, exterior, de la ciencia contemporánea. Apuesto a que podemos ver más y *mejor* si empezamos aquí y ahora, que si intentamos cualquier otro rumbo. Esto no es sólo un prejuicio mío —he estado buscando—sino de que la única manera que conozco de convencerle a usted de que tengo razón es continuar con el proyecto y dejar que los resultados hablen por sí mismos.

La ciencia popular y la imagen manifiesta

¿Qué vemos, pues, cuando miramos este bullicioso mundo público? Alguno de los fenómenos más interesantes y complicados lo ofrecen las acciones de nuestros semejantes. Si tratamos de predecirlas y describirlas utilizando los mismos métodos y conceptos que hemos desarrollado para describir los desprendimientos de tierra, la germinación y el magnetismo, podremos hacer algunos avances, pero el grueso de su macroactividad perceptible —su "comportamiento"— es desesperadamente impredecible a partir de estas perspectivas. La gente es aun más imposible de predecir que el tiempo, si confiamos en las técnicas científicas de los meteorólogos y hasta de los biólogos. Hay, sin embargo, otra perspectiva con la que estamos familiarizados desde la infancia, y que usamos sin esfuerzo todos los días, que parece maravillosamente apta para explicar esta complejidad. Con frecuencia se la llama "ciencia popular". Es la perspectiva que convoca a la familia de conceptos "mentalistas" tales como la fe, el deseo, el conocimiento, el temor, el dolor, la esperanza, la intención, la comprensión, los sueños, la imaginación, la timidez, etcétera.

Se pueden realzar las características importantes de la psicología popular destacando su parecido con otro aspecto de nuestro legado común: la física popular. La física popular es el sistema de expectativas sensatas que todos tenemos acerca de cómo los objetos físicos de tamaño mediano de nuestro mundo reaccionan ante los acontecimientos de mediana importancia. Si vuelco un vaso con agua en la mesa de la cena, usted salta de su silla porque espera que el agua se derrame por el costado y le empape la ropa. Sabe que no puede tratar de absorber el agua con el tenedor, igual que sabe que no puede voltear una casa o empujar una cadena. Espera que un columpio vuelva atrás cuando lo impulsa.

Parte de la física popular puede ser innata, pero por lo menos una parte

es necesario aprenderla. Virtualmente desde la primera infancia, los bebés se encogen de miedo cuando aparece una sombra que les parece amenazadora, y una vez que empiezan a gatear y adquieren la visión estereoscópica (aproximadamente después de los seis meses) se muestran poco dispuestos a aventurarse por encima del "peñasco visual" —una límpida superficie de vidrio que cubre la tabla de la mesa— aun cuando nunca hayan conocido, por amarga experiencia propia, las consecuencias de caerse de un lugar alto (Gibson, 1969). Los niños tienen que aprender, por medio de experiencias individuales, que no pueden caminar por sobre el agua, y que las inestables torres de cubos se vendrán abajo. Parte de la física popular parece estar sustentada por una propensión perceptiva innata: cuando se muestra un dibujo animado de algo que aparentemente está cayendo (por ejemplo, círculos de color "cayendo" como lluvia en una pantalla de vídeo), si se manipula indebidamente el ritmo de aceleración uno ve, en forma instantánea e imposible de reprimir, que alguna fuerza invisible está "empujando" los círculos hacia arriba o abajo para perturbar su movimiento "correcto".

El hecho de que el criterio acerca de la física popular sea innato, o simplemente irresistible, no sería garantía ninguna de su veracidad. La verdad en física académica es, con frecuencia, fuertemente contraintuitiva, o, en otras palabras, contraria a los dictados de la física popular, y no hace falta que descendamos a las perplejidades de la moderna física de partículas en busca de ejemplos. La ingenua física de los líquidos no predecía fenómenos tan sorprendentes y aparentemente mágicos como los sifones y las pipetas (Hayes, 1978), y cualquier persona no iniciada pero inteligente podía deducir con toda facilidad a partir de los claros principios iniciales de la física popular que los giroscopios, las imágenes virtuales producidas por espejos parabólicos, y aun navegar con viento en contra era completamente imposible.

Lo mismo ocurre con la psicología popular. Sus interpretaciones son tan naturales y fáciles que es prácticamente imposible reprimirlas. Imagínese a alguien que está recogiendo arándanos sin tener la menor idea de lo que está haciendo. Imagine ver a dos niños tirando del mismo osito de juguete y que a usted no se le ocurra que ambos lo *quieren*. Cuando un ciego no reacciona ante algo que tiene justo frente a sus ojos, podemos sobresaltarnos, tan apremiante es nuestra expectativa normal de que la gente llega a *creer en la verdad* de lo que ocurre ante sus ojos.

Algunas de las categorías de la psicología popular, como las de la física popular, reciben, en apariencia, un impulso perceptivo innato. Por ejemplo, las pruebas (no concluyentes obtenidas de los estudios en bebés sugieren que la percepción de rostros como categoría preferencial está asistida por mecanismos visuales innatos y algo especializados (Maurer y Barrera, 1981); pero véase también Goren y otros, 1975 y Cohen, DeLoache y Strauss, 1979). A un adulto que no pudiera interpretar un gesto amenazador (o seductor) como tal, se le supondría víctima de una lesión cerebral, no simplemente de haber llevado una vida aislada. Y, sin embargo, todavía hay mucho que debemos aprender en el regazo de nuestra madre, y hasta en la escuela, antes de ser "lectores" expertos del comportamiento de otros en términos mentalísticos (véanse, por ej., Shaftz, Wellman y Silver, 1983; Wimmer y Perner, 1983).

Las intuiciones generadas por la psicología popular no son probablemente más irresistibles, al principio, que las de la física popular, pero quizá debido al estado relativamente no evolucionado y no autoritario de la psicología académica (incluyendo sus parientes cercanas, las neurociencias), hay pocos casos polémicos conocidos en los que la ciencia desacredite en forma directa una intuición de la psicología popular.

¿Cuáles son los sifones y giróscopos de la psicología? Como observa Churchland (1986), "Siempre que el cerebro funcione normalmente, las insuficiencias de la armazón del sentido común pueden ocultarse de la vista, pero con un cerebro dañado se desenmascaran los fallos de la teoría (pág. 223. De modo que debemos observar primero los asombrosos casos anormales. La ceguera (Weiskrantz, 1983) y los fenómenos del cerebro partido (Gazzaniga, 1985) ya han llamado la atención de los filósofos (por ej., Marks, 1980 y Nagel, 1979); están luego la negación de la ceguera y el hemidescuido que el profesor de literatura creyó que yo estaba inventando. (Churchland, 1986, págs. 222-35) ofrece un estudio preliminar. Sacks, 1984, 1986, proporciona descripciones vívidas de algunos casos especialmente extraños, incluyendo su propia experiencia con la "pérdida" temporal de la pierna izquierda). La psicología académica todavía no tiene una teoría oficial acerca de estos fenómenos que pueda oponer a nuestra incredulidad popular, así que siguen siendo polémicas, cuando menos.

Nadie duda de que hay ilusiones perceptivas, y algunas de ellas —por ejemplo, la habitación deformante de Ames (Ittleson, 1952; Gregory, 1977)— enfurecen nuestras expectativas ingenuas. Están además los masoquistas, que tienen fama de disfrutar del dolor (¿?¡!) y la legión de legendarios (y por cierto a veces míticos) sabios idiotas (Smith, 1983). Finalmente está la gente que tiene supuestamente memorias fotográficas, o personalidades múltiples, por no mencionar (y lo digo en serio) a aquellos que tienen supuestos poderes psíquicos.

Este surtido heterogéneo de desafíos a nuestros presentimientos psicológicos diarios debería haber sido suficiente como para volvernos cautelosos al formular reclamos apriorísticos basados en el análisis de los conceptos cotidianos acerca de lo que puede y no puede ocurrir, si bien los filósofos han conferido habitualmente una sorprendente autoridad a dichos conceptos. Tengamos en cuenta los debates filosóficos acerca del autoengaño y la debilidad de la voluntad. Nadie duda de que los fenómenos así llamados por la psicología popular son ubicuos. La polémica consiste en cómo, si es que se puede describir esos fenómenos de forma coherente en términos de fe, conocimiento, intención, criterio y demás términos estándar de la psicología popular, los artículos que llevan títulos como "¿Cómo es posible la falta de voluntad?" (Davidson, 1969) intentan decir exactamente lo que uno debe creer, pensar, saber, intentar y desear para sufrir un estado auténtico de falta de voluntad. Las paradojas y contradicciones que perturban los intentos han desanimado a unos pocos participantes. Para ellos es especialmente claro que las categorías de la psicología popular que aprendieron en la primera infancia son las categorías correctas que hay que usar, cualquiera que sea la perplejidad hageliana que su uso pueda traer en su secuela.

Todos hemos aprendido a adoptar una actitud más escéptica ante los dic-

tados de la física popular, incluyendo los dictámenes sólidos que persisten frente a la ciencia académica. Ni el "hecho introspectivo innegable" de que puede *sentirse* "la fuerza centrífuga" puede evitarla, excepto a los fines pragmáticos de la comprensión rudimentaria a la que siempre ha servido. La delicada pregunta acerca de cómo deberíamos expresar nuestra disminuida fidelidad a las categorías de la física popular ha sido un tema dominante en la filosofía desde el siglo XVII, cuando Descartes, Boyle y otros comenzaron a considerar el status metafísico del color, la sensación del color y las otras "cualidades secundarias". Estas discusiones, aunque cautelosamente agnósticas acerca del status de la física popular, tradicionalmente han adoptado como base indiscutible las categorías complementarias de la psicología popular: las *percepciones conscientes* del color, las *sensaciones* de calor o las *convicciones* acerca del "mundo exterior", por ejemplo. [Esta hipótesis es especialmente evidente en la discusión de Kripke (1972) acerca del materialismo, por ejemplo.]

Algunos de nosotros (Quine, 1960; Dennett, 1969, 1978a; Churchland, 1981; Stich, 1983) nos hemos preguntado si los problemas con lo que nos encontramos en la filosofía tradicional de la mente pueden ser problemas con todo el marco o el sistema de los conceptos de la psicología popular, y hemos recomendado exponerlos al mismo riesgo que a los conceptos de la física popular. No hemos estado de acuerdo con el veredicto, un tópico a ser investigado en los capítulos que siguen, pero sí acerca de la vulnerabilidad, en principio, de los conceptos mentalistas.

La fe que nos sentimos tentados a depositar en las categorías de la psicología popular, tal como nuestra fe en las categorías de la física popular, no se debe únicamente a la obstinada lealtad a la visión del mundo con la que crecimos. En su ensayo clásico *Philosophy and the Scientific Image of Man*, Sellars (1963, capítulo 1) llama a esta visión del mundo la imagen manifiesta, y la diferencia de la imagen científica. No es accidental que tengamos la imagen manifiesta que tenemos; nuestros sistemas nerviosos fueron diseñados como para hacer las diferenciaciones que necesitamos de forma rápida y confiable, como para colocar bajo rótulos sensitivos únicos, las características comunes pertinentes de nuestro entorno, e ignorar todo aquello de lo que habitualmente podamos desentendernos (Dennett, 1984d, por aparecer; Akins, inédito). El hecho innegable es que habitualmente, en especial en los comportamientos más importantes de nuestra vida cotidiana, la ciencia popular funciona. Gracias a la física popular podemos mantenernos abrigados y bien alimentados y evitar los choques, y gracias a la psicología popular colaboramos en proyectos multipersonales, aprendemos los unos de los otros y disfrutamos de períodos de paz local. Estos beneficios serían inalcanzables sin sistemas de expectativa y generación extraordinariamente eficientes y dignos de confianza.

¿Cómo estamos capacitados para hacer todo esto? ¿Qué es lo que organiza nuestra capacidad de tener todas estas expectativas fáciles, seguras y sumamente dignas de confianza? ¿Hay *leyes* o *principios* generales de la física popular que de alguna manera interiorizamos y luego explotamos en forma inconsciente para generar las infinitamente variadas y sensibles expectativas que tenemos acerca de los objetos inanimados? ¿Cómo hacemos para adquirir semejante capacidad *general* para interpretar a nuestros semejantes?

No tengo ninguna explicación que ofrecer acerca de nuestro talento como físicos populares, o acerca de la relación entre la física popular y su prole académica (si bien éste es un tópico fascinante que merece ser mejor estudiado), pero sí tengo una explicación acerc del poder y el éxito de la psicología popular: nos comprendemos los unos a los otros adoptando la actitud intencional.

2

Los verdaderos creyentes: La estrategia intencional y por qué funciona

Habla la Muerte

En Bagdad había un mercader que mandó a su sirviente al mercado a comprar provisiones, y al poco rato el sirviente regresó, blanco y tembloroso y dijo: "Amo, cuando estaba en la plaza del mercado una mujer de la multitud me empujó, y cuando me di vuelta vi que era la Muerte la que me empujaba. Me miró e hizo un gesto amenazador.

"Ahora présteme su caballo y escaparé de esta ciudad para evitar mi destino. Iré a Samarra y allí la Muerte no podrá encontrarme." El mercader le prestó el caballo y el sirviente lo montó, hundió las espuelas en sus flancos y partió todo lo velozmente que el caballo era capaz de galopar. Luego el mercader fue a la plaza del mercado y me vio de pie en medio de la multitud, y se me acercó y me dijo: "¿Por qué le hiciste un gesto amenazador a mi criado cuando lo viste esta mañana?". "Ese no fue un gesto amenazador", le dije. "Fue sólo un respingo de sorpresa. Estaba asombrado de verlo en Bagdad, puesto que yo tenía una cita con él esta noche en Samarra."

W. SOMERSET MAUGHAM

En las ciencias sociales, hablar de *creencias* es inquietante, puesto que los científicos sociales son típicamente tímidos en lo que respecta a sus métodos, hay también mucha palabrería acerca de la *discusión de las creencias*. Y puesto que la creencia es un fenómeno genuinamente extraño, que causa perplejidad y que muestra al mundo muchas caras distintas, hay mucha polémica. A veces el atributo de la fe parece ser un asunto oscuro, arriesgado e imponderable —especialmente cuando creencias exóticas y más especialmente religiosas y supersticiosas están en el candelero. Estos no son los únicos casos conflictivos; también provocamos la discusión y el escepticismo cuando les atribuimos creencias a animales no humanos, a los bebés o a las computadoras o robots. O cuando las creencias que nos sentimos forzados a atribuir a un miembro adulto y aparentemente sano de nuestra propia sociedad son contradictorias, o simplemente de una falsedad feroz. A un biólogo colega mío lo llamó por teléfono una vez en un bar un hombre que quería que él dirimiera una apuesta. El hombre preguntó: "¿Los conejos son pájaros?". "No", dijo el biólogo. "¡Maldición!", exclamó el hombre al colgar. Ahora bien, ¿podría él *realmente* haber creído que los conejos eran pájaros? ¿Se le podría atribuir a alguien verdaderamente esa creencia? Tal vez, pero haría falta un buen cuento para hacer que la aceptáramos.

En todos estos casos la atribución de creencias parece estar acosada por la subjetividad, infectada de relativismo cultural, propensa a la "impresión de la traducción radical", una empresa que evidentemente exige talentos especiales, el arte del análisis fenomenológico, la hermenéutica, la empatía, *Verstehen* y todo eso. En otras ocasiones normales, cuando el tema son las creencias conocidas, la atribución de creencias parece tan fácil como hablar en prosa, y tan objetiva y confiable como contar judías en un plato. Especialmente cuando se nos presentan estos casos directos, es del todo plausible suponer que, en principio (aunque no en la práctica), sería posible confirmar estas atribuciones de creencias simples y objetivas ***encontrando algo dentro de la cabeza del creyente***, en las creencias mismas, en realidad. Alguien podría decir: "Mire... ¿usted cree o no cree que hay leche en el refrigerador?". (En el último caso, usted podría no tener opinión.) Pero si usted sí cree esto, es un hecho perfectamente objetivo acerca de usted, y debe obedecer finalmente a que su cerebro estaba en un estado físico determinado. Si supiéramos más acerca de psicología fisiológica, podríamos en principio determinar el estado de su cerebro y a partir de allí determinar si usted cree o no que hay leche en el refrigerador, aun cuando usted se hubiera decidido a mantenerse en silencio, o en actitud solapada, acerca del tema. En principio, acerca de este punto de vista, la psicología podría superar los resultados —o falta de resultados— de cualquier método de la "caja negra" en las ciencias sociales que conjeture creencias (y otros rasgos mentales) según criterios *externos* de conducta, culturales, sociales, históricos.

Estas reflexiones diferentes convergen en dos creencias opuestas acerca de la naturaleza de la atribución de creencias, por tanto, en la naturaleza de la creencia. Esta última, una variedad del ***realismo***, equipara la pregunta de si alguien tiene determinada creencia en la pregunta de si alguien está infectado por un virus determinado: una cuestión, de hecho, interna, perfectamente objetiva acerca de la cual un observador puede hacer a menudo conjeturas educadas de gran fiabilidad. Lo primero, a lo que podríamos llamar *interpretacionismo* si nos viéramos obligados a ponerle un nombre, equipara la pregunta de si una persona tiene determinada creencia con la pregunta de si una persona es inmoral, o tiene estilo, o talento, o si sería una buena esposa. Enfrentados a semejantes problemas, prologamos nuestras respuestas con "bueno, todo depende de lo que a usted le interese", o admitimos más o menos la relatividad del tema. "Es un caso de interpretación", decimos. Estas dos opiniones contrarias, tan claramente expresadas, no representan en verdad ninguna posición de los teóricos serios, pero sí expresan puntos de vista que se ven típicamente como mutuamente exclusivos y completos en sí mismos. El teórico debe estar a favor de uno y sólo uno de estos puntos.

Creo que esto es un error. Mi tesis será que mientras la creencia es un fenómeno perfectamente objetivo (lo que aparentemente me convierte en un realista), puede ser discernido solamente desde el punto de vista de alguien que adopta cierta ***estrategia predictiva***, y cuya existencia puede ser confirmada sólo por una evaluación del éxito de esa estrategia (lo que aparentemente me convierte en un interpretacionista).

Primero describiré la estrategia, a la que llamo estrategia intencional, o

adoptar la actitud intencional. Para una primera aproximación, la estrategia intencional consiste en tratar al objeto cuyo comportamiento se quiere predecir como un agente racional con creencias y deseos y otras etapas mentales que exhiben lo que Brentano y otros llaman *intencionalidad*. La estrategia ha sido descrita con frecuencia anteriormente, pero trataré de poner este material muy conocido bajo una luz nueva mostrando *cómo* funciona y *cuán bien* lo hace.

Luego sostendré que cualquier objeto —o como lo expresaré, cualquier sistema— cuyo comportamiento esté bien pronosticado por esta estrategia. es, en el más completo sentido de la palabra, un creyente. *Lo que es* ser un veradero creyente es ser un *sistema intencional*, un sistema cuyo comportamiento se puede predecir en forma confiable y amplia por medio de la estrategia intencional. He discutido antes esta postura (Dennett, 1971, 1976b, 1978a) y hasta ahora mis argumentos han reunido pocos conversos y muchos presuntos ejemplos contrarios. Volveré a tratar de hacerlo aquí, más rigurosamente y me ocuparé también de varias objeciones compulsivas.

La estrategia intencional y cómo funciona

Hay muchas estrategias, algunas buenas, otras malas. He aquí una estrategia, por ejemplo, para predecir el futuro comportamiento de una persona: determinar la fecha y hora del nacimiento de la persona y luego alimentar con este dato modesto a uno u otro algoritmo astrológico para generar predicciones acerca de las perspectivas de esa persona. Esta estrategia es deplorablemente popular. Su popularidad es deplorable sólo porque tenemos muy buenas razones para creer que no funciona (Paz Feyerabend, 1978). Cuando las predicciones astrológicas se cumplen no es nada más que pura suerte, o el resultado de una vaguedad o ambigüedad tal en la profecía que casi cualquier eventualidad se puede deducir para confirmarla. Pero supongamos que la estrategia astrológica funcionara bien con cierta gente. Podríamos llamar a esa gente *sistemas astrológicos* —sistemas cuyo comportamiento era, en realidad, predecible mediante la estrategia astrológica—. Si existiera gente así, sistemas astrológicos así, estaríamos mucho más interesados que lo que realmente estamos en *cómo actúa la estrategia astrológica* —es decir, estaríamos interesados en las reglas, principios o métodos de la astrología. Podríamos averiguar cómo actúa la estrategia preguntándoselo a los astrólogos, leyendo sus libros y observándolos en acción. Pero también sentiríamos curiosidad por saber *por qué* funciona. Podríamos descubrir que los astrólogos no tenían ninguna respuesta útil para esta última pregunta —o no tenían teoría alguna de por qué funcionaba, o si sus teorías eran pura palabrería—. Tener una buena estrategia es una cosa; saber por qué funciona es otra.

Por lo que sabemos, sin embargo, la clase de sistemas astrológicos está vacía, lo que quiere decir que la estrategia astrológica interesa solamente como curiosidad social. Otras estrategias tienen credenciales mejores. Tómese en cuenta la estrategia física, o actitud física. Si se quiere predecir el comportamiento de un sistema, determínese su constitución física (quizás hasta ba-

jar el nivel microfísico) y la naturaleza física de los impactos que sufre, y utilícense los conocimientos de las leyes de la física para predecir el resultado para cualquier entrada de datos. Esa es la estrategia más importante, y poco práctica, de Laplace para pronosticar todo el futuro de todo lo que existe en el universo, pero tiene versiones más modestas, locales y realmente practicables. El químico o el físico puede utilizar esta estrategia en el laboratorio para predecir el comportamiento de materiales exóticos, pero también la cocinera que está en la cocina, puede predecir el efecto de dejar la olla demasaido tiempo sobre el fuego. Esa estrategia no es siempre viable en la práctica, pero que *en principio* siempre funcionará es un dogma de las ciencias físicas. (Desconozco las complicaciones menores provocadas por las imprecisiones subatómicas de la física cuántica.)

De cualquier modo, a veces es más eficaz cambiar de la actitud física a lo que llamo la actitud de diseño, donde uno desconoce los detalles reales (probablemente complicados de la constitución física de un objeto, y, sobre la suposición de que tiene cierto diseño, predice que se comportará *como está diseñado para comportarse* en distintas circunstancias. Por ejemplo, la mayoría de quienes usan ordenadores, no tienen la menor idea de qué principios físicos son los responsables del comportamiento altamente fiable y por lo tanto fácil de pronosticar, del ordenador. Pero si tienen alguna idea de para qué está diseñado (una descripción de su funcionamiento en cualquiera de los muchos niveles de abstracción posibles), podrían predecir su comportamiento con gran exactitud y fiabilidad, sujeto a no ser confirmado sólo en los casos de mal funcionamiento físico. Menos dramáticamente, casi todos pueden predecir cuándo sonará un reloj despertador sobre la base de la inspección más casual de su exterior. Uno no sabe ni le importa saber si es a cuerda, a pila, a energía solar, si tiene un mecanismo de bronce y rubíes a chips de silicio —uno simplemente da por sentado que está diseñado como para que la alarma suene a la hora que se puso, que esa hora está bien indicada, y que el reloj seguirá andando hasta esa hora y más allá de ella, y que está para funcionar más o menos exactamente y así sucesivamente. Para obtener pronósticos más detallados y exactos acerca del diseño del reloj despertador, se debe descender a un nivel de descripción más abstracto de su diseño, por ejemplo al nivel en que se describen los engranajes, pero sin especificar de qué material están hechos.

Naturalmente, el comportamiento de un sistema proyectado sólo es predecible a partir de lo intentado en el diseño. Si se desea predecir el comportamiento de un reloj despertador que se lo ha llenado de helio líquido, hay que volver a la actitud física. No sólo los artefactos sino también muchos seres vivos (plantas y animales, riñones y corazones, estambres y pistilos) se comportan en forma tal que se puede predecir por cómo están formados. No son sólo sistemas físicos sino sistemas diseñados.

A veces hasta la actitud de diseño es prácticamente innacesible y entonces aun hay otra estrategia o actitud más que se puede adoptar: la actitud intencional. He aquí cómo funciona: primero se decide tratar al objeto cuyo funcionamiento hay que predecir como un agente racional; luego se deduce qué creencias debería tener ese agente, dada su posición en el mundo y su objetivo. Más tarde se deduce qué deseos tendría que tener siguiendo las mis-

mas consideraciones, y por fin se predice que este agente racional actuará para conseguir sus metas a la luz de sus creencias.

En muchos, pero no en todos los casos, un poco de razonamiento práctico a partir del conjunto de creencias o deseos elegidos, producirá una decisión acerca de lo que el agente debería hacer; eso es lo que se predice que el agente *hará*.

La estrategia se hace algo más clara con un poco de explicación. Tenga en cuenta primero cómo vamos de un lado a otro llenándonos las cabezas unos a otros con creencias. Algunos axiomas: la gente que se aísla tiende a ser ignorante; si alguien se expone a algo acaba por saberlo todo a su respecto. Parece que, en general, llegamos a creer todas las verdades acerca de las partes del mundo que nos rodea si estamos colocados en una posición adecuada como para aprender. La exposición a x, es decir, la confrontación sensorial con x durante un lapso adecuado, es la condición *normalmente suficiente* para saber (o tener creencias verdaderas) acerca de x. Como decimos, llegamos a *saber todo acerca* de las cosas que nos rodean. Esa exposición es sólo *normalmente* suficiente para el conocimiento, pero no es la gran escotilla de escape que parece ser; nuestro umbral para aceptar la ignorancia anormal cuando se hace frente a la exposición es muy alto. "Yo no sabía que el arma estaba cargada", dicho por alguien a quien se le vio presente, consciente y despierto cuando se cargó el arma, choca con un escepticismo de todo tipo que sólo la historia de apoyo más extravagante podría superar.

Por supuesto, no llegamos a aprender o recordar todas las verdades que nuestras historias sensoriales nos ofrecen. A pesar de la frase "saber todo acerca de", lo que llegamos a saber, normalmente, son solamente las verdades *pertinentes* que nuestras historias sensoriales nos sirven. No alcanzo yo a saber la relación que hay entre la gente con lentes y la que usa pantalones en una habitación en la que vivo, aunque si me interesa de verdad, sería muy fácil saberlo. No se trata simplemente de que algunos hechos de mi entorno estén por debajo de mi umbral de discriminación o más allá del poder de integración y retención de mi memoria (como, por ej., la estatura en pulgadas de todas las personas presentes) sino de que muchos hechos perfectamente detectables, comprensibles y recordables carecen de interés para mí y por lo tanto no llego a creer en ellos. De manera que una regla para atribuir creencias en la estrategia intencional es ésta: atribuya como creencias todas las verdades pertinentes al interés (o los deseos) que la experiencia del sistema hasta este momento haya hecho asequibles. Esta regla lleva a presuponer demasiado, pues todos somos algo olvidadizos, incluso con las cosas importantes. Tampoco logra captar las creencias falsas que se sabe que todos tenemos, pero la atribución de una creencia falsa, *cualquier* creencia falsa, exige una genealogía especial que, como se verá, consiste, en su mayor parte, en creencias verdaderas. Dos casos paradigmáticos: S cree (falsamente) que p, porque S cree (de verdad) que Jones le dijo que p, que Jones es muy inteligente, que Jones no tenía intenciones de engañarle... etcétera. Segundo caso: S cree (falsamente) que hay una víbora en el taburete del bar porque S cree (de verdad) que le parece ver una víbora en el taburete y él mismo está sentado ante el mostrador a poco menos de un metro de distancia del taburete que ve, y así sucesivamente. La falsedad tiene que empezar en alguna parte;

se puede sembrar la semilla en estado de alucinación, engaño, una variedad normal de percepción falsa, deterioro de la memoria o fraude deliberado, por ejemplo, pero las creencias falsas que se cosechan crecen en un medio de cultivo de creencias verdaderas.

Luego están las creencias arcanas y sofisticadas, verdaderas y falsas, que tan a menudo son el centro de atención en las discusiones acerca de la atribución de creencias. Dios sabe que éstas no surgen directamente de la exposición a cosas y hechos mundanos, sino que su atribución exige descubrir una serie de mayormente muy buenos argumentos o razonamientos en el montón, de creencias ya atribuidas. Por tanto, una deducción de la estrategia intencional es que los verdaderos creyentes creen principalmente en verdades. Si alguien lograra idear un método acordado para individualizar y contar creencias (mucho lo dudo) veríamos que todas, excepto la menor parte (digamos, menos del diez por ciento) de las creencias de una persona fueron atribuibles según nuestra primera regla. [1]

Téngase en cuenta que esta regla es una regla derivada, una elaboración y ulterior especificación de la regla fundamental: atribúyanse aquellas creencias que el sistema *tendría que tener.* Obsérvese también que la regla interactúa con la atribución de los deseos. ¿Cómo atribuimos los deseos, preferencias, objetivos, intereses, sobre cuyas bases daremos forma a la lista de creencias?

Atribuimos los deseos que el sistema *tendría que tener.* Esa es la regla fundamental. Ella dictamina, primero, que atribuyamos a la gente la lista conocida de los deseos más básicos y elevados: supervivencia, ausencia de dolor, comida, comodidades, procreación, entretenimiento. Citar cualquiera de estos deseos termina con el juego de "¿Por qués?", de explicar razones. No se supone que necesitemos un motivo ulterior para desear la comodidad o el placer, o la prolongación de nuestra existencia, las reglas derivadas de la atribución de deseos interactúan con las atribuciones de creencias. En forma trivial, tenemos la regla: atribúyanse los deseos de aquellas cosas que un sistema considera buenos para sí. De manera algo más informativa, atribuya los deseos de cosas que un sistema crea que son el mejor medio para lograr

[1] *La idea de que la mayoría de las creencias de alguien debe* ser verdadera le parece evidente a mucha gente. Se puede buscar el apoyo a esta idea en las obras de Quine, Putnam, Shoemaker, Davidson, y en ls mías. Otra gente considera la idea igualmente increíble, de manera que es posible que cada parte esté llamando creencia a un fenómeno distinto. Una vez que se hace la distinción entre creencia y opinión (en mi sentido técnico —véase *How to Change your Mind* en *Brainstorms*, capítulo 16—) según el cual las opiniones están contaminadas desde el punto de vista lingüístico, los estados cognitivos relativamente sofisticados —que son aproximadamente estados de apostar a la verdad de una afirmación formulada especial— se puede ver la casi trivialidad del alegato de que la mayoría de las creencias son verdaderas. Unas pocas reflexiones acerca de temas periféricos lo tendría que sacar a relucir. Téngase en cuenta a Demócrito, que tenía una física sistemática, abarcadora, pero (digamos, por el *bien* de la discusión) que era completamente falsa. Estaba del todo equivocado, aunque sus puntos de vista se mantenían y tenían una especie de utilidad sistemática. Pero aun si todas las afirmaciones que la erudición nos permite atribuir a Demócrito (implícitas o explícitas en sus escritos) son falsas, representan una fracción casi inexistente de sus *creencias*, que incluyen tanto al gran número de creencias permanentes y vulgares que debe de haber tenido (acerca de la casa en que vivía, lo que buscaba en un buen par de sandalias, y así sucesivamente) y también esas creencias ocasionales que iban y venían por millones a medida que su experiencia preceptiva cambiaba.

otros fines a los que aspira. La atribución de deseos estrafalarios y perjudiciales, exige, por tanto, como la atribución de noticias falsas, relatos especiales.

La relación entre creencia y deseo se hace más engañosa cuando consideramos qué deseos atribuimos sobre la base de la conducta verbal. La capacidad de *expresar* deseos mediante el habla abre las compuertas de la atribución de deseos. "Quiero una tortilla de setas de dos huevos, pan francés y mantequilla y media botella de Borgoña blanco bien fresco." ¿Cómo se puede comenzar a atribuir un deseo de algo tan específico sin semejante declaración verbal? En realidad, ¿cómo podría alguien *contraer* un deseo tan específico sin la ayuda del lenguaje? El lenguaje *nos permite* formular deseos muy específicos pero también *nos obliga*, en ocasiones, a comprometernos con deseos cuyas condiciones de satisfacción son a la postre más severas que cualquier otra cosa que de otro modo tendríamos alguna razón en esforzarnos por satisfacer. Puesto que para obtener lo que se quiere a menudo hay que decirlo, y puesto que con frecuencia no se puede decir lo que se quiere sin decir algo más específico que lo que se quiso decir antes, a veces se termina dándoles a otros pruebas —la mejor prueba, nuestra palabra extorsionada— de que deseamos cosas o estados de cosas mucho más especiales que las que nos satisfarían —o mejor aun— que nos habrían satisfecho, puesto que una vez que uno lo ha declarado, por ser una persona de palabra, uno adquiere interés en satisfacer exactamente ese deseo que declaró y ningún otro.

"Por favor, quiero habas al horno."

"Sí señor. ¿Cuántas?"

Al exigírsenos semejante especificación de un deseo, podríamos muy bien oponernos, si bien, en realidad, todos estamos suficientemente socializados como para acceder a exigencias semejantes en la vida cotidiana —hasta el punto de no darnos cuenta y, por cierto, de no sentirnos oprimidos por ella. Me extiendo en este punto porque tiene un paralelo en el reino de las creencias, en el que nuestro entorno linguístico siempre nos está obligando a dar —o conceder— una expresión verbal precisa a conviciones a las que les fal-

---

Pero, se puede hacer notar, este aislamiento de sus creencias vulgares de su ciencia depende de una distinción insostenible entre las verdades de la observación y las de la teoría: todas las creencias de Demócrito están recargadas de teoría, y, puesto que esta teoría es falsa, todas lo son. La respuesta es la siguiente: concedido que todas las creencias de observación están recargadas de teoría, ¿por qué deberíamos elegir la teoría *explícita* y sofisticada de Demócrito (expresada en sus *opiniones*) como la teoría con la que recargar sus observaciones cotidianas? Nótese que el compatriota menos teórico de Demócrito también tenía miles de creencias de observación recargadas de teoría —y no era, en este sentido— más sabio por ello. ¿Por qué no suponemos que las observaciones de Demócrito están recargadas con la misma (presuntamente inocua) teoría? Si Demócrito olvidó esta teoría, o cambió de idea, sus creencis de observación estarían intactas *en gran parte*, hasta el punto de que su sofisticada teoría jugó un papel evidente en su conducta rutinaria, en sus expectativas y demás, sería sumamente adecuado expresar sus creencias vulgares desde el punto de vista de la teoría sofisticada, pero esto no produciría un catálogo *mayormente falso* de creencias, puesto que sólo algunas de éstas resultarían afectadas. (A menudo se subestima el efecto de la teoría sobre la observación. Véase Churchland, 1979, para encontrar ejemplos interesantes y convincentes de la estrecha relación que puede existir a veces entre la teoría y la experiencia.) [La discusión que aparece en esta nota fue extraída de una útil conversación con Paul y Patricia Churchland y Michael Stack.]

tan los estrechos límites con los que los dota la verbalización (véase Dennett 1969, págs. 184-85, y *Brainstorms*, capítulo 16). Al concentrarse en los *resultados* de esta fuerza social, sin tener en cuenta su efecto deformante, se puede llegar a pensar, equivocadamente, que es *evidente* que las creencias y los deseos son como oraciones almacenadas en la cabeza. Al ser criaturas parlantes, es inevitable que, a menudo, lleguemos a creer que una oración determinada, realmente formulada, deletreada y puntuada es *verdadera*, y que en otras ocasiones lleguemos a querer que esa oración *se haga realidad*, pero éstos son casos especiales de creencia y deseo y como tales pueden no ser modelos fiables para la totalidad del campo de acción.

Esto es suficiente, en esta ocasión, sobre los principios de atribución de deseos y creencias que se encuentran en la estrategia intencional. ¿Y qué hay de la racionalidad que se le atribuye a un sistema intencional? Se empieza con el ideal de racionalidad perfecta y se revisan según lo dicten las circunstancias. Es decir, se empieza por la suposición de que la gente cree en todas las implicaciones de sus creencias, y no cree en dos pares contradictorios de creencias. Esto no crea un problema práctico de desorden (infinitas implicaciones, por ejemplo), puesto que sólo se está interesado en asegurar que el sistema que se está prediciendo es lo bastante racional como para llegar a las implicaciones particulares que son inherentes a su predicamento de conducta en ese momento. Los casos de irracionalidad o de capacidad de deducción limitadamente poderosas, crean problemas de interpretación especialmente intrincados, que descartaré en esta ocasión (véase capítulo 4, *Making Sense of Ourselves*, y Cherniak, 1986).

Quiero pasar de la descripción de la estrategia al tema de su uso. ¿La gente utiliza en verdad esta estrategia? Sí, siempre. Algún día podrán haber otras estrategias para atribuir creencias y deseos y para predecir la conducta, pero ésta es la única que conocemos ahora. ¿Y cuándo? La gente actúa así siempre. ¿Por qué *no* sería buena idea permitir que las diferentes facultades de Oxford creen y otorguen rangos académicos cada vez que lo consideran adecuado? La respuesta es una larga historia, muy fácil de imaginar, y habría un amplio consenso acerca de los puntos más importantes. No tenemos ninguna dificultad para adivinar en las razones que la gente tendría entonces para actuar de manera tal como para darles a otros motivos para.... creando así circunstancias indeseadas. Nuestra utilización de la estrategia intencional es tan común y fácil que el papel que desempeña cuando da forma a nuestras expectativas acerca de la gente se ignora con facilidad. La estrategia funciona también en la mayoría de los mamíferos casi siempre. Por ejemplo, se puede usar para diseñar mejores trampas para cazar esos mamíferos, razonando sobre lo que la criatura sabe o cree acerca de distintas cosas, qué prefiere, qué quiere evitar. La estrategia funciona con los pájaros, con los peces, con los reptiles y con los insectos y arañas y hasta con criaturas tan inferiores y poco emprendedoras como las almejas. (Cuando una almeja cree que hay algún peligro cerca, no afloja su apretón sobre su concha cerrada hasta que se convence de que ha pasado el peligro.) También funciona con algunos artefactos: el ordenador que juega al ajedrez no se comerá mi caballo porque sabe que hay una línea de juego siguiente que le llevaría a

perder su torre, y no quiere que eso suceda. Más modestamente el termostato apagará la caldera en cuanto llegue a creer que la habitación ha alcanzado la temperatura deseada.

La estrategia funciona hasta con las plantas. En un lugar con tormentas tardías de primavera, habría que plantar variedades de manzana que son especialmente *cautelosas* en lo que se refiere a *llegar a la conclusión* de que es primavera, que es cuando quieren florecer, por supuesto. También funciona con fenómenos tan inanimados y aparentemente no intencionales como el rayo. Una vez un electricista me explicó cómo logró proteger la bomba de agua de mi sótano del daño producido por los rayos; el rayo, me dijo, siempre busca encontrar el mejor camino para llegar a la tierra, pero a veces se engaña y toma atajos no tan buenos como el primero. Se puede proteger la bomba trazando otro atajo mejor y más claro para el rayo.

Los verdaderos creyentes como Sistemas Intencionales

Así éste es un surtido heterogéneo de atribuciones "serias" de creencias, atribuciones dudosas de creencias, metáforas pedagógicamente útiles, *façons de parler* y, quizá peor, fraudes cabales. La siguiente tarea sería distinguir aquellos sistemas intencionales que ***realmente*** tienen creencias y deseos. Pero ese sería un trabajo de Sísifo, o, de lo contrario, se terminaría por decreto. Una comprensión mejor del fenómeno de la creencia comienza por la observación de que hasta en el peor de estos casos, aun cuando estuviéramos completamente seguros que la estrategia funciona *por razones equivocadas*, es sin embargo cierto que sí funciona, por lo menos un poquito. Este es un hecho interesante, que diferencia esta clase de objetos, la clase de *sistemas intencionales*, de la clase de objetos para los que la estrategia nunca funciona. Pero ¿es así? ¿Nuestra definición de un sistema intencional excluye algún objeto? Por ejemplo parece que el atril de este salón de conferencias puede ser interpretado como sistema intencional, totalmente racional, creyendo que está colocado actualmente en el centro del mundo civilizado (como quizá lo piensen también algunos de ustedes), y deseando sobre todo permanecer en ese centro. ¿Qué tendría que hacer ese agente racional así equipado con creencias y deseos? Evidentemente, quedarse quieto, que es exactamente lo que el atril hace. Predigo el comportamiento del atril exactamente, desde la actitud intencional, por tanto, ¿es éste un sistema intencional? Si lo es, absolutamente cualquier cosa lo es.

¿Qué descalificaría al atril? Por un lado, en este caso la estrategia no se recominda a sí misma, puesto que no obtenemos de ella ningún poder predictivo que ya no tuviéramos anteriormente. Ya sabíamos lo que el atril iba a hacer —o sea nada— y adaptamos las creencias y los deseos para que se ajusten de una manera sumamente desprovista de principios. Sin embargo en el caso de las personas, los animales o los ordenadores, la situación es diferente. En estos casos, con frecuencia la única estrategia práctica es la estrategia intencional; nos brinda un poder predictivo que no podemos obtener por ningún otro método. Pero se debe insistir que esto no supone ninguna diferencia

en la esencia, sino simplemente una diferencia que se refleja en nuestra limitada capacidad como científicos. El omnisciente físico laplaciano podría predecir el comportamiento de un ordenador —o de un cuerpo humano vivo, dando por sentado que está gobernado finalmente por las leyes de la física— sin necesidad de los métodos arriesgados, rápidos y directos de las estrategias intencionales o diseñadas. Para la gente que tiene una aptitud mecánica limitada, la interpretación intencional de un simple termostato es una muleta práctica y sumamente inocua, pero los ingenieros que hay entre nosotros pueden entender del todo su funcionamiento interno sin la ayuda de esta antropomorfización. Puede ser verdad que a los ingenieros más inteligentes les resulta prácticamente imposible conservar una concepción clara de sistemas más complejos, tales como el sistema computarizado de tiempo compartido o una sonda espacial a control remoto, sin caer en una actitud intencional (y considerando estos artefactos como algo que pregunta y relata, prueba y evita, quiere y cree), pero éste sólo es un caso más avanzado de debilidad epistemológico-humana. No quisiéramos clasificar estos artefactos junto a los verdaderos creyentes —nosotros mismos— sobre bases tan variables y parroquiales, ¿no?, ¿no sería intolerable juzgar que algún artefacto o ente o persona era un creyente desde el punto de vista de un observador, pero para nada un creyente desde el punto de vista de otro observador más inteligente? Esa sería una versión especialmente radical del interpretacionismo, y algunos han creído que yo la adoptaba al insistir en que la creencia fuera vista en los términos del éxito de la estrategia intencional. Debo confesar que mi presentación de ese punto de vista a veces ha invitado a esta lectura, pero ahora quiero restarle fuerza. La decisión de adoptar la actitud intencional es libre, pero los hechos acerca del éxito o del fracaso de la actitud, si se la adoptara, son perfectamente objetivos.

Una vez que la estrategia intencional está en su lugar, es una herramienta extraordinariamente poderosa para la predicción, un hecho que está ampliamente ocultado por nuestra típica concentración en los casos en los que produce resultados dudosos o no confiables. Tenga en cuenta, por ejemplo, la predicción de los movimientos en un partido de ajedrez; lo que hace del ajedrez un juego interesante es la impredictibilidad de los movimientos del rival, excepto en aquellos casos en que los movimientos son "forzados" —donde *claramente* hay un movimiento mejor— típicamente el menor de los males posibles. Pero esta impredictibilidad se ubica en el contexto cuando uno reconoce que en la situación ajedrecística tipo hay muchísimos movimientos perfectamente legales y por tanto disponibles, pero sólo unos pocos —tal vez media docena— que sean algo recomendable, y de ahí que de acuerdo con la estrategia intencional hay sólo unos pocos movimientos de alta probabilidad. Aun cuando la estrategia intencional no logre distinguir un único movimiento con las mayores probabilidades, puede reducir drásticamente el número de opciones de interés.

La misma característica es aparente cuando se aplica la estrategia intencional a casos del "mundo real". Es obvio, incapaz de predecir las decisiones exactas de compra y venta de quienes juegan a la bolsa, por ejemplo, o la exacta secuencia de palabras que un político pronunciará al decir un discurso programado, pero nuestra confianza puede por cierto ser muy alta en lo

que respecta a predicciones ligeramente menos específicas: que ese jugador de bolsa hoy *no comprará acciones de una empresa pública*, o que el político *se pondrá de parte de los sindicatos en contra de su partido*, por ejemplo. Esta incapacidad de predecir descripciones de acciones precisas, miradas de otra manera, es una fuente de fortaleza para la estrategia intencional puesto que esta neutralidad respecto de los detalles de la práctica, es lo que permite explotar la estrategia intencional en casos complejos, por ejemplo, en *encadenar predicciones* (véase *Brainstorms*). Supongamos que el Secretario de Estado de los Estados Unidos anunciara que era agente de la KGB. ¡Qué acontecimiento incomparable! ¡Cuán impredecibles sus consecuencias! Sin embargo lo cierto es que podemos predecir docenas de consecuencias no terriblemente interesantes pero sí sobresalientes. El Presidente deliberaría con el resto del Gabinete que apoyaría su decisión de relevar al Secretario de Estado de sus obligaciones hasta conocer los resultados de distintas investigaciones psiquiátricas y políticas, y de todo esto se informaría en una conferencia de prensa a gente que escribiría cuentos que serían comentados en editoriales que serían leídos por gente que les escribiría cartas a los editores, y así sucesivamente. Nada de esto es una predicción temeraria, pero obsérvese que describe un arco de causalidad en el espacio-tiempo que no podría predecirse bajo *ninguna* descripción por ninguna extensión práctica imaginable de la física o la biología.

El poder de la estrategia intencional se puede ver aun más exactamente con la ayuda de una objeción planteada por primera vez por Robert Nozick hace algunos años. Supongamos, sugirió, que algunos seres de inteligencia muy superior —de Marte digamos— descendieran sobre nosotros, y supongamos que para ellos nosotros fuéramos como los termostatos para ingenieros inteligentes. Es decir, supongamos que no *necesitaran* la actitud intencional —ni siquiera actitud de diseño— para predecir nuestra conducta en todos sus detalles. Se puede suponer que sean super-físicos laplaceanos capaces de entender la actividad de Wall Street, por ejemplo, a nivel microfísico. Allí donde vemos edificios y agentes de bolsa y órdenes de venta y licitaciones, ellos verían un cúmulo de partículas subatómicas arremolinándose— y son tan buenos físicos que podrían predecir con días de anticipación qué marcas de tinta aparecerán cada día en la cinta de papel rotulada "El promedio industrial de cierre de Dow Jones". Podrían predecir las conductas individuales de los distintos cuerpos en movimiento que observan sin siquiera tratar a ninguno de ellos como sistemas intencionales. ¿Tendríamos entonces razón en decir que desde *su* punto de vista no éramos realmente creyentes en absoluto (no más de lo que lo es un simple termostato)? Si así fuera, nuestro status de creyentes no es nada objetivo, sino más bien algo que el espectador ve —siempre que el espectador comparta nuestras limitaciones intelectuales.

Nuestros marcianos imaginarios podrían predecir el futuro de la raza humana por medio de métodos laplaceanos, pero si no nos vieran también como sistemas intencionales, estarían omitiendo algo perfectamente objetivo: los *modelos* del comportamiento humano que se pueden describir desde la actitud intencional y sólo desde esa actitud, y que sustentan la generalizaciones y las predicciones. Tomemos un caso particular en el que los mar-

cianos observan a un corredor de bolsa colocar una orden de 500 acciones de la General Motors. Predicen los movimientos exactos de sus dedos al marcar el número de teléfono y las vibraciones exactas de sus cuerdas vocales al hacer su pedido. Pero los marcianos no ven que una gran cantidad de modelos *diferentes* de movimientos de los dedos y vibraciones de las cuerdas vocales —hasta los movimientos de muchísimos individuos diferentes— podrían haber sido sustituidos por los detalles reales sin alterar la operación siguiente del mercado, entonces habrían dejado de ver un modelo real en el mundo que están observando. Así como hay cantidades infinitas de maneras de *ser una bujía* —y uno no ha entendido lo que es un motor de combustión interna a menos que se dé cuenta de que una gran variedad de dispositivos diferentes se pueden atornillar en estas cavidades sin afectar el rendimiento del motor—, también hay infinitas maneras de *ordenar 500 acciones de la General Motors* y hay cavidades sociales en las cuales una de estas maneras producirá casi el mismo efecto que cualquier otra. También hay puntos pivotes sociales, por así decirlo, donde para qué lado va la gente depende de si creen que *p*, o desean *A* y no depende de ninguna de las otras infinitas maneras que pueden ser semejantes o diferentes.

Supongamos, llevando nuestra fantasía marciana un poco más lejos, que uno de los marcianos participara en un concurso de predicciones con un terrícola. El terrícola y el marciano observan (y observan como el otro observa) una pizca determinada de una transacción física local. Desde el punto de vista del terrícola esto es lo que se observa. Suena el teléfono en la cocina de la Sra. Gardner. Ella contesta y esto es lo que dice: "Hola querido, ¿volverás a casa temprano? ¿Dentro de una hora? ¿Y traes al jefe a cenar? Compra entonces una botella de vino por el camino y conduce con cuidado". A partir de esta observación, nuestro terrícola predice que un vehículo metálico grande con llantas de goma se detendrá en el sendero de la casa en una hora y descargará dos seres humanos, uno de los cuales llevará una bolsa de papel que contiene una botella con un líquido alcohólico. La predicción es tal vez un poco arriesgada, pero es en verdad una buena apuesta. El marciano hace la misma predicción pero tiene que buscarse mucha más información acerca de una cantidad extraordinaria de interacciones de las que hasta donde él sabe, el terrícola desconoce completamente. Por ejemplo, la desaceleración del vehículo en la intersección *A*, a unos ocho kilómetros de la casa, sin la cual habría habido un choque con otro vehículo, cuya trayectoria de choque había sido arduamente calculada por el marciano en algunos cientos de metros. La actuación del terrícola ¡parecería magia! ¿Cómo sabía el terrícola que el ser humano que bajó del auto y compró la botella en la tienda volvería a subir? El hecho de que la predicción del terrícola se hiciera realidad, después de todas las extravagancias, cruces y ramificaciones en los senderos trazados por el marciano, le parecería a cualquier persona desprovista de la estrategia intencional, algo tan maravilloso e inexplicable como la inevitabilidad fatalista de la cita en Samarra. Los fatalistas —por ejemplo los astrólogos— creen que hay un modelo inexorable en los asuntos humanos, que se impondrá *venga lo que...*, es decir, no importa cómo las víctimas proyecten, qué justificaciones inventen, no importa cómo se retuerzan y den vueltas en sus cadenas. Estos fatalistas están equivocados, pero están *casi* en lo cierto. *Hay* mo-

delos en los asuntos humanos que se imponen, no de manera completamente inexorable pero con gran vigor, absorbiendo perturbaciones y variaciones físicas que también podrían ser consideradas fortuitas; estos son los modelos que caracterizan en los términos de las creencias, deseos e intenciones de los agentes racionales.

Sin duda usted habrá notado, y se habrá sentido aturdido por una falta grave en nuestro experimento acerca del pensamiento: se supone que el marciano trata a su oponente terrícola como un ser tan inteligente como él, con quien es posible la comunicación, un ser con el que se puede hacer una apuesta, contra quien se puede competir. En pocas palabras, un ser con creencias (tales como la que manifestó en su predicción) y deseos (tales como el deseo de ganar el concurso de predicciones). Por tanto, si el marciano ve el modelo en un terrícola, ¿cómo es que no lo ve en los demás? Como una pequeña parte de un relato, nuestro ejemplo podría fortalecerse suponiendo que nuestro terrícola tuvo la inteligencia de aprender el idioma marciano (que se transmite por modulación de rayos X) y se disfrazó de marciano, contando con que el chauvinismo de esos extraños, por otra parte muy brillantes, le permitiría pasar por sistema intencional sin delatar, al mismo tiempo, el secreto de sus congéneres humanos. Añadir esto podría permitirnos superar un mal giro de la historia, pero podría oscurecer la moraleja de la que sacar en conclusión: es decir, ***la inevitabilidad de la actitud intencional con referencia a uno mismo y sus congéneres inteligentes.*** Esta inevitabilidad es de interés relativo por sí misma; es perfectamente posible adoptar una actitud física, por ejemplo, acerca de un ser inteligente, incluido uno mismo pero no hasta el punto de excluir la posibilidad de mantener, al mismo tiempo, una actitud intencional al menos acerca de uno mismo y sus semejantes *si* uno intenta, por ejemplo, enterarse de lo que saben (punto en el que hizo hincapié enérgicamente Stuart Hampshire en muchos escritos). Tal vez podemos suponer que nuestros superinteligentes marcianos no nos reconozcan como sistemas intencionales, pero no podemos suponer que carezcan de los conceptos necesarios. [2] Si ellos conservan, teorizan, predicen, se comunican; se ven a ellos mismos como sistemas intencionales. [3] Allí donde hay seres inteligentes, debe haber modelos que los describan, nos interese, o no, verlos.

Es importante reconocer la realidad objetiva de los modelos intencionales discernibles en las actividades de los seres inteligentes, pero es también

[2] Un integrante del auditorio en Oxford señaló que si el marciano incluía al terrícola en su campo de acción de actitud física (posibilidad que yo, explícitamente, no había excluido), él no se sentiría sorprendido por la predicción del terrícola. Habría predicho, en efecto, exactamente el diseño de las modulaciones de rayos X producidas por el terrícola hablando el idioma marciano. Es verdad, pero cuando el marciano anotara el resultado de sus cálculos, aparecería su predicción de la predicción del terrícola, palabra por palabra marciana, como en una pizarra espiritista, y que lo que sería desconcertante para el marciano sería cómo ese trozo de mecanismo, el pronosticador disfrazado de marciano, era capaz de producir esta oración *real* en marciano estando aislado informativamente de los hechos que el marciano necesitaba conocer para hacer su propia predicción acerca del automóvil que llegaba.

[3] ¿Podría no haber seres inteligentes para quienes la comunicación, la predicción y la observación fueran inútiles...? Podrían haber innumerables entidades excelentes y maravillosas que carecieran de estos modos de acción, pero no veo qué nos llevaría a llamarlas *inteligentes.*

importante reconocer lo incompleto y las imperfecciones de los modelos. El hecho objetivo es que la estrategia intencional *trabaja tan bien como puede*, que no es a la perfección. Nadie es perfectamente racional, nadie tiene una memoria perfecta, nadie es un observador perfecto o invulnerable a la fatiga, al mal funcionamiento o a la imperfección del diseño. Esto lleva inevitablemente a circunstancias que están más allá del poder de descripción de la estrategia intencional, del mismo modo en que el daño físico a un artefacto, como un teléfono o un automóvil, lo puede volver indescriptible por la terminología normal del diseño de dicho artefacto. ¿Cómo se describe el diagrama esquemático de la instalación alámbrica de un amplificador de audio que se ha difundido parcialmente? o ¿cómo se explica el estado del programa de un ordenador que funciona mal? En los casos de hasta la más leve y conocida patología cognitiva —cuando la gente parece tener creencias contradictorias o estar engañándose a sí mismo, por ejemplo— los cánones de interpretación de la estrategia intencional no logran producir veredictos claros y constantes acerca de qué creencias y deseos atribuirle a una persona.

Ahora bien, una posición realista *fuerte* acerca de las creencias y los deseos alegaría que en estos casos la persona en cuestión tiene verdaderamente deseos y creencias determinados que la estrategia intencional, tal como la he descrito, es simplemente incapaz de adivinar. En la clase más benigna de realismo que estoy defendiendo, no existe evidencia de exactamente qué creencias y deseos tiene una persona en estos casos degenerados, pero esto no significa entregarse al relativismo o subjetivismo, porque *cuándo* y *por qué* no hay evidencia verdadera es en sí misma una cuestión objetiva. Según este punto de vista, se puede hasta reconocer *la relatividad del interés* de la atribución de creencias y conceder que, dados los distintos intereses de las diferentes culturas, por ejemplo, las creencias y deseos que una cultura atribuiría a uno de sus miembros, podrían ser muy distintos de las creencias y deseos que otra cultura le atribuiría a esa misma persona. Pero suponiendo que eso fuera así en un caso determinado, estarían los hechos ulteriores acerca de *cuán* bien las estrategias intencionales rivales funcionaban para predecir la conducta de dicha persona. Podemos estar seguros por adelantado que ninguna interpretación intencional de un individuo funcionará a la perfección, y que puede ser que dos esquemas rivales sean igualmente buenos, y mejores que otros que podemos idear. Que esto sea así es por sí mismo algo que puede ser la verdad del caso. La presencia objetiva de un modelo (cualquiera que sean sus imperfecciones) no excluye la presencia objetiva de otro modelo (cualquiera que sean sus imperfecciones).

El fantasma de interpretaciones radicalmente diferentes con idéntica garantía desde la estrategia intencional es teóricamente importante —sería mejor decir metafísicamente importante— pero prácticamente insignificante una vez que uno limita su atención a los sistemas intencionales mayores y más complejos que conocemos: los seres humanos. [4]

Hasta ahora he estado destacando nuestra afinidad con las almejas y los termostatos para enfatizar un punto de vista del status lógico de la atribu-

[4] La analogía de John McCarthy con la criptografía se esmera en esto con delicadeza. Cuanto mayor sea el cuerpo del texto cifrado, menor probabilidad hay de que se lo descifre en forma

ción de creencias pero ha llegado el momento de reconocer las diferencias evidentes y decir qué conclusión sacar. El reclamo perverso perdura: para ser un verdadero creyente *no hay más que* ser un sistema cuya conducta se pueda predecir de manera fiable por medio de la estrategia intencional, y por lo tanto *no hay más que* creer real y verdaderamente que *p* (para cualquier proposición *p*) es un sistema intencional para el cual *p* ocurre como creencia en la interpretación mejor (más predecible). Pero una vez que volcamos nuestra atención a los sistemas intencionales verdaderamente interesantes y versátiles, vemos que este criterio de creencia instrumentalista aparentemente superficial reprime severamente la constitución interna de un creyente genuino y así produce después de todo una versión robusta de la creencia.

Considérese el termostato modesto, un caso tan degenerado de sistema intencional que no mantedrá de manera plausible nuestra atención más de un momento. Siguiendo con la fantochada, podríamos ponernos de acuerdo en concederle la capacidad de sólo una media docena de creencias distintas y aun de menos deseos: puede creer que la habitación es demasiado fría o demasiado calurosa, que la caldera está prendida o apagada y que si quiere que la habitación esté más abrigada debe encender la caldera, y así sucesivamente. Pero seguramente esto sería imputarle demasiado al termostato; no tiene idea del calor ni de la caldera, por ejemplo. Por lo tanto supongamos que *desinterpretemos* sus creencias y deseos: puede creer que *A* es también *F* o *G* y si quiere que *A* sea más *F* tendría que hacer *K*, y así sucesivamente. Después de todo, agregando el mecanismo de control termostático a los diferentes dispositivos de entrada y salida, se le podría hacer regular la cantidad de agua de un tanque, o la velocidad de un tren, por ejemplo. Su unión a un transductor sensible al calor y a una caldera es una conexión con el mundo demasiado empobrecida como para garantizarles cualquier semántica rica a sus estados semejantes a creencias.

Pero supóngase que luego enriquecemos estas formas de unión. Supongamos que se le proporciona más de una manera de enterarse de la temperatura, por ejemplo. Le proveemos de una especie de ojo que distinga a los acurrucados y temblorosos ocupantes de la habitación y un oído para poder decirle cuánto frío hace. Le damos algunos datos geográficos que le permitan llegar a la conclusión de que está probablemente en un lugar frío si se entera de que su posición espacio-temporal es Winnipeg en diciembre. Por supuesto, el otorgarle un sistema visual que es multipropósito y general —no un mero detector de objetos temblorosos— exigirá grandes complicaciones en su estructura interior. Supongamos que también le otorgamos a nuestros sistemas una mayor versatilidad de conducta: elige el combustible para la caldera. lo compra en el concesionario más barato y fiable, revisa los burletes, etc. Esto agrega otra dimensión de complejidad interna; les da a los estados indivi-

---

dual y sistemáticamente inconexa. Véase McCarthy, 1979, para una discusión muy útil de los principios y presuposiciones de la actitud intencional aplicada a los aparatos —incluyendo explícitamente a los termostatos.

duales semejantes a creencias *más quehacer*, en efecto, al proporcionarle más y diferentes oportunidades para su derivación o deducción de otros estados y al proporcionarle más y diferentes oportunidades para que ellas sirvan como premisas para razonamientos ulteriores. El efecto acumulativo de enriquecer estas conexiones entre el dispositivo y el mundo en el que reside es enriquecer la semántica de sus falsos predicados, *F*, *G* y el resto. Cuantos más agregamos menos dispuesto se vuelve nuestro dispositivo a servir como estructura de control de otra cosa que no sea un sistema de mantenimiento de la temperatura de una habitación. Una manera más formal de expresar esto es que la clase de modelos satisfactorios no distinguibles del sistema formal, incorporados a sus estados internos, se reduce y reduce a medida que agregamos tales complejidades. Cuantas más agregamos, más rica o más exigente o más específica se vuelve la semántica del sistema, hasta que eventualmente lleguemos a los sistemas para los cuales se ha dictado prácticamente (pero nunca en principio) una interpretación semántica única (véase Hayes, 1979). En ese punto decimos que este dispositivo (o animal o persona) tiene creencias ***acerca del calor y acerca de esta habitación***, y así sucesivamente, no sólo debido a la ubicación real del sistema y su funcionamiento en el mundo, sino también porque no podemos imaginarnos otro lugar adecuado en el que se pudiera colocar y ***en el que funcionara*** (véase también los capítulos 5 y 8).

Nuestro simple termostato original tenía un estado que llamábamos creencia acerca de determinada caldera, en cuanto a si estaba encendida o apagada. ¿Por qué a cerca de ***esa*** caldera? Pues bien, ¿acerca de qué otra caldera quisiera usted que fuera? La creencia es acerca de la caldera porque está ***fijado*** a ella. [5] Dada la conexión causal real, si bien mínima, con el mundo, que estaba en vigor, podríamos dotar a un estado del dispositivo de ***significado*** (o algo parecido) y ***condiciones de verdad***, pero era demasiado fácil sustituir una conexión mínima diferente y cambiar completamente el significado (en este sentido empobrecido) de ese estado interno. Pero a medida que los sistemas se vuelven más ricos desde el punto de vista perceptivo y más versátiles desde el punto de vista de la conducta, se hace más y más difícil hacer cambios en las conexiones verdaderas del sistema con el mundo sin cambiar la organización del sistema mismo. Si se modifica su entorno, ***lo notará***, en efecto, y producirá un cambio en su estado interno como respuesta. Ocurre una represión de doble sentido en el carácter específico del crecimiento entre el dispositivo y el entorno. Fijemos el dispositivo en algún otro estado y exigiremos un entorno muy específico en el cual funcionar en forma adecuada (ya no se puede cambiarlo fácilmente de regular la temperatura a regular la velocidad o cualquier otra cosa); pero al mismo tiempo si no ***fijamos*** el estado en el que está sino que lo tiramos en un entorno cambiado, sus fijaciones sensoriales serán lo suficientemente sensibles y discriminatorios como para responder adecuadamente al cambio, llevando el sistema a un nuevo estado en el cual funcionará con eficiencia en el nuevo entorno. Hay una manera conoci-

[5] Esta idea es en realidad la predecesora de las especies de ideas diferentes amontonadas bajo la rúbrica de una creencia de *re*. Si uno parte de esta idea para llegar a sus retoños, se pueden ver mejor las dificultades con ellos, y cómo repararlas. (Véase el capítulo 5. ***Más allá de la creencia***, para más información acerca de este tema).

da de referirse a esta estrecha relación que puede existir entre la organización de un sistema y su entorno: se dice que el organismo *refleja* continuamente el entorno, o que hay una *representación* del entorno en la organización del sistema, o implícita en ella.

No se trata de que atribuyamos (o debamos atribuir) creencias y deseos sólo a las cosas en las que encontramos representaciones internas, sino que cuando descubrimos algún objeto para el cual la estrategia intencional funciona, nos esforzamos por interpretar algunos de sus estados o procesos internos como representaciones internas. Lo que hace que algún rasgo interno de algo sea una representación sólo podría ser su papel en la regulación de un sistema intencional.

Ahora debiera estar clara la razón por la que recalcamos nuestra afinidad con el termostato. No existe ningún momento mágico en la transición de un termostato simple a un sistema que tiene *realmente* una representación interna del mundo que lo rodea. El termostato tiene una representación mínimamente exigente del mundo. Los termostatos más finos tienen representaciones más exigentes del mundo; los robots más complejos para ayudar en la casa tendrían representaciones del mundo aun más exigentes. Por fin se llega a nosotros. Estamos conectados al mundo de manera tan múltiple e intrincada que casi no hay sustitución posible; aunque es claramente imaginable en un experimento acerca del pensamiento. Hilary Putnam imagina el Planeta Tierra Gemelo que es exactamente igual a la Tierra hasta en las marcas de las caminatas en los zapatos de la copia del Planeta Tierra Gemelo que tiene su vecino, pero que se diferencia de la Tierra en cierta propiedad que está completamente por debajo de los umbrales de nuestra capacidad de discriminación (lo que llaman agua en el Tierra Gemelo tiene un análisis químico diferente). Si a *usted* se lo enviara instantáneamente al Tierra Gemelo y se lo cambiara por su réplica, usted jamás sería más sabio, tal y como el sistema simple de control que no sabe distinguir si está regulando la temperatura, la velocidad o el volumen de agua de un tanque. Es fácil inventar Planetas Tierras Gemelos radicalmente distintos para algo tan sencillo y tan desprovisto sensorialmente como un termostato, pero su organización interna tiene una exigencia mucho más rigurosa para la sustitución. Sus Planeta Tierra Gemelo y Tierra deben ser copias casi exactas o usted cambiará dramáticamente de estado al llegar.

¿Entonces acerca de qué caldera son *sus* creencias cuando usted cree que la caldera está encendida? Pues de la caldera que está en su sótano (más que de su gemela en el Planeta Tierra Gemelo, por ejemplo). ¿Acerca de qué otra caldera podrían ser sus creencias? La terminación de la interpretación semántica de sus creencias, fijando los referentes de éstas, exige, tal como en el caso del termostato, hechos reales acerca de su verdadera inserción en el mundo. Los principios y problemas, de interpretación que descubrimos cuando atribuimos creencias a la gente son los *mismos* principios y problemas que descubrimos cuando miramos el problema ridículo, pero gloriosamente sencillo, de atribuirle creencias a un termostato. Las diferencias son

de categoría, pero obstante, de una categoría tan grande que comprender la organización interna de un sistema intencional simple nos da muy poca base para entender la organización interna de un sistema intencional complejo, tal como un ser humano.

¿Por qué funciona la estrategia intencional?

Cuando volvemos a la pregunta de *por qué* la estrategia intencional funciona tan bien, descubrimos que la pregunta es ambigua y que admite dos tipos de respuestas muy diferentes. Si el sistema intencional es un termostato simple, una respuesta es sencillamente ésta: la estrategia intencional funciona porque el termostato está bien diseñado. Fue diseñado para ser un sistema que podía ser comprendido y manipulado fiable y fácilmente desde esta actitud. Eso es cierto, pero no muy informativo, si lo que se busca son las características reales de su diseño que expliquen su funcionamiento. Por fortuna, sin embargo, en el caso de un termostato simple esas características se encuentran y se entienden con facilidad, de manera que la otra respuesta a nuestra pregunta de *por qué*, que es en realidad una respuesta acerca de *cómo funciona el mecanismo*, está a mano sin demora.

Si el sistema intencional en cuestión es una persona, también hay una ambigüedad en nuestra pregunta. La primera respuesta a la pregunta de por qué funciona la estrategia intencional es que la evolución ha diseñado a los seres humanos para ser racionales, para creer lo que deben creer y desear lo que deben desear. El hecho de que seamos los productos de un proceso evolutivo largo y exigente garantiza el hecho de que usar la estrategia intencional en nosotros sea una apuesta segura. Esta respuesta tiene la virtud de la veracidad y la brevedad, y en esta ocasión la virtud adicional de ser una respuesta que Herbert Spencer aplaudiría, pero también es llamativamente poco informativa. La versión más difícil de la pregunta inquiere, en efecto, cómo funciona el mecanismo con el cual la naturaleza nos ha dotado. Y todavía no podemos dar una buena respuesta a esa pregunta. Sencillamente no lo sabemos. *Sí* sabemos cómo funciona la *estrategia*, y conocemos la respuesta fácil a la pregunta de por qué funciona, pero saber esto no nos es muy útil con la respuesta difícil.

Sin embargo, no se trata de que haya una carencia de doctrina. Un conductista skinneriano, por ejemplo, diría que la estrategia funciona porque sus imputaciones de creencias y deseos son, en efecto, versiones taquigráficas de descripciones hasta ahora inimaginablemente complejas de los efectos de historias previas de respuesta y refuerzo. Decir que alguien quiere helado es decir que en el pasado la ingestión de helado fue reforzada en él por los resultados, creando una propensión en ciertas condiciones ambientales (también demasiado complejas de describir) para adquirir una conducta de compra de helados. En ausencia de un conocimiento detallado de esos hechos históricos podemos, sin embargo, hacer deducciones sagaces sobre fundamentos inductivos. Estas deducciones están comprendidas en nuestros reclamos de actitud intencional. Aun si todo esto fuera cierto, nos diría muy poco acerca del modo en que tales propensiones fueron reguladas por el mecanismo interno.

Una explicación, actualmente la más popular, es que el informe acerca de cómo funciona la estrategia y de cómo funciona el mecanismo *coincidirán* (aproximadamente): para cada creencia predictivamente atribuible habrá un estado interno funcionalmente notable del mecanismo, que se puede descomponer en partes funcionales de casi la misma forma en que la oración que expresa la creencia se puede descomponer en parte, es decir, en palabras o términos. Las inferencias que les atribuimos a los seres racionales serán reflejadas por procesos causales físicos en el hardware; la forma *lógica* de las proposiciones en que se cree, se copiarán en la forma *estructural* de los estados en correspondencia con ellos. Esta es la hipótesis de que hay un *lenguaje del pensamiento* codificado en nuestros cerebros, y que eventualmente nuestros cerebros serán entendidos como sistemas manipuladores de símbolos en una *analogía* por lo menos aproximada con los ordenadores. Se están explorando en la actualidad muchas versiones diferentes de este punto de vista, en el nuevo programa de investigación llamado ciencia cognitiva, y siempre que uno permita una gran amplitud para atenuar el audaz reclamo básico, creo que alguna versión de esto demostrará ser correcta.

Pero no creo que esto sea *evidente*. Quienes creen que es evidente o inevitable que esa teoría se demostrará que es verdadera (y hay muchos que lo creen), están confundiendo dos reclamos empíricos diferentes. El primero es que la descripción de la actitud intencional produce como resultado un modelo verdadero y objetivo en el mundo: el modelo que nuestros marcianos imaginarios no percibieron. Este es un reclamo empírico, pero que está confirmado más allá del escepticismo. El segundo es que este modelo verdadero es *producido por* otro modelo real aproximadamente isomórfico a *él* dentro de los cerebros de seres inteligentes. Dudar de la existencia del segundo modelo real no es dudar de la existencia del primero. *Hay* razones para creer en el segundo, pero no son abrumadoras. La mejor explicación simple que puedo dar de las razones es la siguiente.

A medida que ascendemos en la escala de complejidad del termostato simple pasando por el robot sofisticado hasta el ser humano, descubrimos que nuestros esfuerzos por diseñar sistemas con la conducta requerida chocan con el problema de la *explosión combinatoria*. Aumentando algunos parámetros en, digamos, un diez por ciento —un diez por ciento más de datos o más grados de libertad en la conducta a ser controlada o más palabras a ser reconocidas o lo que sea— tiende a aumentar la complejidad interna del sistema diseñado por orden de magnitud. Las cosas escapan muy rápidamente del control y, por ejemplo, pueden conducir a programas de computación que arruinan las máquinas más grandes y veloces. Ahora de algún modo el cerebro ha resuelto el problema de la *explosión combinatoria*. Es una red gigantesca de miles de millones de células, pero todavía limitado, compacto, fiable y veloz y capaz de aprender nuevas conductas, vocabularios, teorías casi sin límite. Algunos principios de representación elegantes, fecundos, infinitamente extensibles deben ser los responsables. Tenemos un solo modelo de semejante sistema de representación: un lenguaje humano. De manera que el argumento a favor de un lenguaje del pensamiento llega hasta esto: ¿Qué otra cosa podría ser? Hasta ahora no hemos sido capaces de imaginar ninguna alternativa plausible en ningún detalle. Creo que ésa es una buena

razón para recomendar como un tema de táctica científica que sigamos la hipótesis en sus distintas formas lo más lejos que podamos. [6] Pero nos enfrascaremos en esa exploración en una forma más circunspecta y fructífera si tenemos en cuenta que su inevitable precisión está lejos de estar asegurada. No se entiende bien ni siquiera una hipótesis verdaderamente empírica mientras uno esté bajo el error de que es necesariamente cierta.

Reflexiones: Modelos reales, hechos más profundos y preguntas vacías

Varias de las cuestiones que reciben un tratamiento rápido en "Los verdaderos creyentes" se extienden hasta ser temas centrales en capítulos subsiguientes. Tal vez la mayor fuente de inquietud acerca de mi posición a través de los años, ha sido su delicado equilibrio acerca del punto de la relatividad del observador de atribuciones de creencia y otros estados intencionales. Desde mi punto de vista, ¿está la atribución (o el significado) de la creencia en el juicio del observador? ¿Creo que hay *verdades subjetivas* acerca de lo que cree la gente, o alego que todas las atribuciones no son más que ficciones útiles? Mi discusión de la objeción de Nozick intenta colocar firmemente mi punto de vista en el filo de la navaja entre los intolerables extremos del realismo simple y el relativismo simple pero ésta no ha sido reconocida como una opción constante y atractiva por muchas otras personas de este campo, y mis críticos han tratado, de manera persistente, de demostrar que mi posición se cae en un abismo u otro.

Insisto en que mi punto de vista es una *especie* de realismo, puesto que sostengo que los modelos que los marcianos no perciben están real y objetivamente allí para ser notados o pasados por alto. ¿Cómo pudieron los marcianos, que lo "saben todo" acerca de los hechos físicos de nuestro mundo no percibir estos modelos? ¿Qué significaría decir que algunos modelos, aunque estén objetivamente allí, son visibles sólo desde un punto de vista? Un elegante micromundo de dos dimensiones brinda un ejemplo claro: "el Juego de la vida" de John Horton Conway (Gardner, 1970), una fuente simple pero extraordinariamente rica de instrospecciones que deberían hacerse parte de los recursos imaginativos de todos como un lugar de ensayo versátil para los experimentos sobre el pensamiento acerca de la relación de niveles en la ciencia.

Imagine un pedazo grande de papel cuadriculado. Las intersecciones (no los cuadrados) de esta red son los únicos lugares —llamados celdas— en el micromundo de la Vida, y en cualquier instante cada célula está Encendida o Apagada. Cada célula tiene ocho vecinos: las cuatro células adyacentes

[6] El hecho de que todos los modelos de representación mental del *lenguaje del pensamiento* propuesto hasta ahora caigan víctimas de la explosión combinatoria debería de una u otra manera atemperar el entusiasmo por participar en lo que Fodor llama adecuadamente "el único juego del pueblo".

al norte, sur, este y oeste de ella y las cuatro células diagonales más próximas (noreste, sudeste, sudoeste, noroeste). En el mundo de la Vida el tiempo también es discreto, no continuo; avanza a tictacs, y el estado del mundo cambia entre cada tictac de acuerdo con la siguiente regla:

> Para determinar qué hacer en el instante siguiente, cada célula cuenta cuántos de sus ocho vecinos están Encendidos en ese instante. Si la respuesta es exactamente dos, la célula permanece en su estado actual (Encendida o Apagada) en el instante siguiente. Si la respuesta es exactamente tres, la célula se Enciende en el instante siguiente cualquiera que sea su estado actual. En todas las demás condiciones la célula está apagada.

Toda la física determinista del mundo de la Vida se atrapa en esa única ley común y corriente. (Así como ésta es la ley fundamental de la "física" del mundo de la Vida, ayuda al principio a concebir esta física extraña en términos biológicos: piense en las células Encendiéndose como nacimientos, Apagándose como muertes, y en los instantes subsiguientes como generaciones. Tanto el exceso de gente como el aislamiento causan la muerte; el nacimiento ocurre sólo en circunstancias propicias.) Mediante la aplicación escrupulosa de la ley, se puede predecir con perfecta exactitud el momento siguiente de cualquier forma de las células Encendidas y Apagadas, y el instante posterior a ése, y así sucesivamente. De manera que el mundo de la Vida es un excelente corralito laplaceano: un mundo determinista simplificado en el cual nosotros, seres limitados, podemos adoptar la actitud física y predecir el futuro con suprema confianza. Existen muchas simulaciones computadorizadas del mundo de la Vida, en las que se pueden poner configuraciones en pantalla y luego mirarlas evolucionar según la norma. En las mejores simulaciones, se puede cambiar tanto la escala del tiempo como la del espacio, alternando entre el primer plano y la vista a vuelo de pájaro.

Pronto se descubre que algunas configuraciones simples son más interesantes que otras. Hay cosas que titilan de un lado para otro entre dos configuraciones, cosas que crecen y luego se desintegran, "pistolas planeadoras" que emiten "planeadores" —configuraciones que se reproducen a sí mismas traducidas unas pocas células más adelante, deslizándose gradualmente a través del paisaje bidimensional— locomotoras de vapor, bombarderos, comedores al paso, anticuerpos, rastrillos espaciales. Una vez que uno comprende la conducta de estas configuraciones, puede adoptar la actitud de diseño y preguntarse cómo diseñar grandes conjuntos de estos objetos que realizarán tareas más complicadas. Uno de los triunfos de la actitud de diseño en el mundo de la Vida es una máquina de Turing universal: una configuración cuyo comportamiento se puede interpretar como el cambio de estado, la lectura de símbolos y la escritura de símbolos de un ordenador simple y que se puede "programar" para computarizar cualquier función computarizable. (Véase "The Abilities of Men and Men Machines" en *Brainstorms*, y Dennett 1985a para más información acerca de las máquinas de Turing.)

Cualquiera que formule la hipótesis de que alguna configuración del mundo de la Vida es tal que una máquina de Turing puede predecir su estado futuro con precisión, eficiencia, y una pizca de riesgo, debe adoptar la

"actitud de la máquina de Turing" y, pasando por alto tanto la física del mundo de la Vida como los detalles de diseño de la máquina, calcular sólo la función que está siendo computarizada por la máquina Turing; luego volver a traducir el rendimiento de la función al sistema de símbolos de la máquina del mundo de la Vida. Se puede predecir que pronto aparecerá la configuración de Encendidos y Apagados, siempre que ningún planeador extraviado u otros desechos ruidosos choquen con la máquina de Turing y la destruyan o la descompongan.

¿Es "real" el modelo que le permite formular esta predicción? Mientras dure, lo es, y si el modelo incluye una "coraza" para aislar la máquina del ruido, el modelo podrá sobrevivir por algún tiempo. El modelo *puede* deberle su existencia a las intenciones (perspicaces o confusas) del diseñador de la máquina, pero su realidad en cualquier sentido interesante —su longevidad o vigor— es estrictamente independiente de los hechos históricos acerca de su origen.

Si el modelo se puede ver o no, es otra cosa. En un sentido sería visible para cualquiera que observara el despliegue de las configuraciones especiales en el plano de la Vida. Una de las delicias del mundo de la Vida es que no hay nada escondido en él; no hay bambalinas. Pero hace falta un gran salto de introspección para ver este despliegue *como* la computarización de una máquina de Turing. Subir hasta el punto de vista desde el que este nivel de explicación y predicción es visible sin dificultad, es opcional, y podría ser difícil para muchos.

Sostengo que la actitud intencional proporciona una posición aventajada para discernir modelos igualmente útiles. Estos modelos son objetivos —están *allí* para ser detectados— pero desde nuestro punto de vista no están *allí* completamente independientes de nosotros, puesto que son modelos compuestos en parte por nuestras propias reacciones "subjetivas" ante lo que está ahí; son los modelos hechos a medida para nuestros intereses narcisistas (Atkins, 1986). Es fácil para nosotros, así constituidos, percibir los patrones que son visibles desde la actitud intencional; y sólo desde esa actitud. [7] A los marcianos les resultaría muy difícil, pero pueden aspirar a conocer las regularidades que son una segunda naturaleza para nosotros tal como nosotros podemos aspirar a conocer el mundo de la araña o del pez.

De manera que soy una especie de realista. Rehúso la invitación para unirme al perspectivismo radical de Rorty (1979, 1982) (Dennett 1982a). Pero afirmo también que cuando estos modelos objetivos no son todo lo perfectos que siempre deberían ser, habrá brechas imposibles de interpretar; en principio siempre es posible para las interpretaciones rivales de la actitud intencional de esos modelos, empatar el primer puesto de manera que ningún

[7] Ascombe habló misteriosamente de "una orden que está allí cada vez que las acciones son ejecutadas con intenciones (1957, pág. 80) pero no dijo *en qué lugar* del mundo se percibía. ¿En el cerebro? ¿En la conducta? Durante años esto no tuvo sentido para mí, pero ahora veo lo que ella puede haber querido decir y por qué fue tan evasiva en su descripción (y ubicación) de la orden. Es tan difícil decir dónde está la orden intencional como decir dónde están los modelos intencionales en el mundo de la Vida. Si se "mira" de la manera correcta, los modelos son evidentes. Si se mira (o describe) en el mundo de cualquier otra manera, son invisibles en general.

hecho ulterior pueda determinar qué creía *realmente* el sistema intencional en cuestión.

Esta idea no es nueva. Es una prolongación completamente directa de la tesis de Quine (1960) acerca de la imprecisión de la tradución radical, aplicada a la "traducción" no sólo de los modelos en la disposición de los sujetos para entregarse a la conducta externa, ("los significados estimulantes "de Quine"), sino también los modelos ulteriores dispuestos a "conducirse") internamente. Como dice Quine, "la traducción radical empieza en casa" (1969, pág. 46) y las implicaciones de su tesis se extienden más allá de su periferialismo o conductismo.

> La metáfora de la caja negra, a menudo tan útil, puede resultar engañosa aquí. El problema no es uno de los hechos ocultos tales como los que podrían destaparse aprendiendo más acerca de la fisiología cerebral de los procesos del pensamiento. Esperar un mecanismo físico distintivo detrás de todo estado mental genuinamente diferente es una cosa; esperar un mecanismo distinto para toda diferencia supuesta que se puede formular en el idioma mentalista tradicional, es otra. La cuestión de si... el extranjero cree *realmente* a *A* o cree más bien a *B*, es una cuestión cuya significación misma pondría en duda. Esto es lo que estoy queriendo decir al discutir la imprecisión de la traducción (Quine 1970, págs. 180-81).

Mi argumento en "Brain Writing and Mind Reading" (1975) ampliado acerca de este asunto, exponiendo el error de aquellos que habían esperado *encontrar algo en el cerebro* para resolver los casos que el periferialismo de Quine dejó irresolutos son exactamente las mismas consideraciones aplicadas a la traducción de cualquier "idioma del pensamiento" que se podría descubrir si se abandonara el conductismo por la ciencia cognitiva. Otro quineano[8] que ha defendido esta posición acerca de la creencia es Davidson (1974a):

> La imprecisión del significado de la traducción no significa el fracaso en captar diferencias significativas. Señala el hecho de que ciertas diferencias aparentes no son significativas. Si hay imprecisión es porque cuando está toda la evidencia, quedan abiertas maneras alternativas de manifestar los hechos. (pág. 332).

Otra explicación de la idea de la imprecisión ha sido muy bien defendida más recientemente por Parfit (1984), que alegó que los principios en los que confiamos (o deberíamos confiar) para decidir cuestiones de *identidad personal*, también dejarán inevitablemente abierta la posibilidad en princi-

[8] Algunos no-quineanos han sostenido versiones de esta idea. Wheeler (1986) demuestra con perspicacia que Derrida "proporciona argumentos y consideraciones suplementarias importantes, si bien peligrosas" a aquellos que han sido superados por Davidson y otros quineanos. Como advierte Wheeler:

> Por supuesto para los quineanos ya es evidente que el habla y el pensamiento son escrituras cerebrales, algún tipo de signos tan sujetos a la interpretación como cualquier otro... sin embargo, en los círculos no-quineanos parece haber una creencia encubierta de que en alguna forma el habla interna expresa directamente el pensamiento (pág. 492).

Esta creencia encubierta se expone al ataque en Capítulo 8.

pio de casos enigmáticos imprecisos. A mucha gente le resulta extraordinariamente difícil aceptar esto.

> La mayor parte de nosotros nos inclinamos a creer que en cualquier caso imaginable, la pregunta "¿me estoy muriendo?"debe tener una respuesta. Y nos sentimos inclinados a creer que esta respuesta debe ser simplemente "Sí" o "No". Cualquier persona futura debe ser o yo o algún otro. Llamo a estas creencias el punto de vista de que *nuestra identidad debe estar definida* (pág. 214).

Nos sentimos tan fuertemente atraídos por la opinión de que *los contenidos de nuestros pensamientos o creencias deben estar definidos*, y resistimos la sugerencia de que la pregunta "¿creo que hay leche en el refrigerador?" puede no tener una respuesta definida "sí" o "no". Parfit demuestra que hay otros casos en los que nos conformamos sin una respuesta a esas preguntas.

> Suponga que cierto club haya existido durante varios años, realizando reuniones regulares. Luego las reuniones cesan. Algunos años más tarde algunos de los socios de este club forman otro con el mismo nombre y los mismos reglamentos. "¿Ha formado esta gente exactamente el mismo club? ¿O han abierto simplemente *otro* club exactamente igual?". Podría haber una respuesta para esta pregunta. El club original podría haber tenido una disposición explicando cómo, después de semejante período de inexistencia se lo podía volver a constituir. O podría haber tenido una disposición que impidiera esto. Pero supongamos que tal regla no exista, como tampoco ningún hecho legal que sustentara cualquiera de las dos respuestas a nuestra pregunta. Y suponga que la gente involucrada, si se les formulara la pregunta, no la responderían. No habría entonces respuesta a nuestra pregunta. La afirmación "este es el mismo club" no sería *ni verdadera ni falsa*. Aunque no hay respuesta a nuestra pregunta, puede no haber nada que no sepamos... Cuando esto ocurre con alguna pregunta yo la llamo una pregunta *vacía* (pág. 213).

Reconocemos que en el caso de la existencia del club y en casos similares, no hay ningún "hecho" que resolvería el punto, pero en el caso de la identidad personal, la suposición de que hay —debe haber— tal hecho más profundo tarda en desaparecer. No es sorprendente que ésta sea una de las convicciones brutales que Nagel (1986) no puede abandonar:

> ¿Por qué no es suficiente identificarme a mí mismo como una persona en el sentido más débil en el cual ésta sea el sujeto de predicados mentales pero no algo con una existencia separada, más como una nación que como un yo cartesiano?
>
> En realidad no tengo una respuesta para esto, excepto la respuesta de que es una pregunta implorante... (pág. 45).

Formulo el caso análogo para lo que podríamos llamar identidad de la creencia o precisión de la creencia. En el Capítulo 8 "Evolution, Error, and Intentionality", muestro como Searle, Fodor, Dretske, Kripke y Burge (entre otros) caen en la tentación de buscar este hecho más profundo, y cómo su búsqueda no tiene esperanza. Sostengo en otras palabras, que alguna de las cuestiones más vigorosamente discutidas acerca de la atribución de creencias son, en el sentido de Parfit, preguntas vacías.

¿Cómo se crea esta persistente ilusión? En las reflexiones acerca de "Making Sense of Ourselves", describo un falso contraste (entre nuestras creencias y aquéllas de los animales inferiores) que crea la convicción errónea de que nuestras creencias y otros estados mentales deben tener un contenido definido.

Otro *leitmotif* que consigue su primer papel en "Los verdaderos creyentes" es la comparación entre el experimento sobre el pensamiento en el Planeta Tierra Gemelo de Putnam y distintos trastornos o intercambios ambientales más sencillos: de los termostatos en este caso. Se desarrolla el tema más plenamente en el capítulo 5 "Más allá de la creencia", y en el ejemplo del "bitser doble" del capítulo 8.

Finalmente, el respaldo atenuado y condicional a la idea de un lenguaje del pensamiento al final de "Los verdaderos creyentes" se elabora con más cuidado en los capítulos 5 y 6, "Estilos de representación mental" y en las reflexiones que lo siguen.

3

Tres clases de psicología intencional*

La psicología popular como fuente de teorías

Suponga que tanto usted como yo creemos que los gatos comen pescado. ¿Exactamente qué rasgo tenemos en común para que esto sea verdad para ambos? De manera más general, y recordando el estilo de pregunta preferido por Sócrates, ¿qué tiene que haber en común entre las cosas a las que realmente se les atribuye un predicado *intencional*, tal como "quiere visitar la China" o "espera fideos para la cena"? Como señala Sócrates en el *Meno* y en otras partes, esas preguntas son ambiguas o vagas en su intención. Se puede estar preguntando por un lado por algo semejante a una definición, o por otro por algo semejante a una teoría. (Por supuesto, Sócrates prefería la primera clase de respuesta). ¿Qué tienen en común todos los imanes? Primera respuesta: todos atraen el hierro. Segunda respuesta: todos tiene tal y tal propiedad microfísica (una propiedad que explica su capacidad para atraer el hierro). En un sentido la gente sabía lo que eran los imanes —cosas que atraían el hierro— mucho antes de que la ciencia les dijera qué eran los imanes. Un niño aprende lo que la palabra "imán" significa, típicamente no aprendiendo una definición explícita, sino aprendiendo la "física popular" de los imanes, en la que el término común "imán" está incorporado o definido implícitamente como término teórico.

A veces los términos están incorporados a teorías muy poderosas, y a veces están incorporados mediante una definición explícita. ¿Qué tienen en común los elementos químicos con la misma valencia? Primera respuesta: están dispuestos a combinarse con otros elementos en las mismas proporciones integrales. Segunda respuesta: todos tienen tal y cual propiedad microfísica (una propiedad que explica su capacidad para combinarse de esa manera). La teoría de las valencias en química estaba disponible antes de que se conociera su explicación microfísica. En algún sentido los químicos sabían qué eran las valencias antes de que los físicos se lo dijeran.

De manera que, lo que en Platón parece un contraste entre dar una definición y dar una teoría se puede observar simplemente como un caso especial del contraste entre dar una respuesta teórica y dar otra repuesta teórica más "reductiva". Fodor traza el mismo contraste entre las respuestas "concep-

* Presentado originalmente ante el Thyssen Philosophy Group y el Bristol Fulbright Workshop, en setiembre de 1978 y reimpreso con la autorización de R. Healy ed. *Reduction, time and Reality* (Cambridge: Cambridge University Press, 1981).

tuales" y "causales" a tales preguntas y afirma que Ryle (1914) defiende las respuestas conceptuales a expensas de las causales, suponiendo erróneamente que están en conflicto. La acusación de Fodor contra Ryle es justa, pues hay, por cierto, muchas partes en las que Ryle parece proponer sus respuestas conceptuales como baluartes contra la posibilidad de preguntas causales, científicas, psicológicas, pero hay una opinión mejor de Ryle (o tal vez, en el mejor de los casos una opinión que él debería haber sostenido) que merece rehabilitación. El "conductismo lógico" de Ryle está compuesto por sus respuestas resueltamente conceptuales a las preguntas socráticas acerca de temas mentales. Si Ryle pensó que estas respuestas descartaban la psicología, que descartaban las respuestas causales (o reductivas) a las preguntas socráticas, estaba equivocado, pero si sólo pensaba que las respuestas conceptuales a las preguntas no fueron dadas por una psicología microrreductiva, pisaba terreno más firme. Una cosa es dar una explicación causal de algún fenómeno y otra completamente distinta citar la causa del fenómeno en el análisis del concepto de éste.

Algunos conceptos tienen lo que podría llamarse un elemento causal esencial (véase Fodor 1975, pág. 7, N° 6). Por ejemplo, el concepto de un *autógrafo* auténtico de Winston Churchill sostiene que la manera en que se trazó la línea de tinta es esencial para su status como autógrafo. Se descartan las fotocopias, falsificaciones, las firmas inadvertidamente imposibles de distinguir, pero quizá no las copias hechas con papel carbón. Estas consideraciones son parte de la respuesta *conceptual* a la pregunta socrática acerca de los autógrafos.

Algunos, incluyendo a Fodor, han afirmado que conceptos tales como el concepto de la acción inteligente, también tienen un elemento causal esencial; se puede demostrar que la conducta que parecía ser inteligente no lo es demostrando que tiene la clase incorrecta de causa. En contra de tales posiciones Ryle puede argumentar que aún si es cierto que todo ejemplo de la conducta inteligente es causado (y tiene por lo tanto una explicación causal), exactamente el *cómo* se causa no es esencial para que sea inteligente; algo que podría ser verdad aún si toda la conducta inteligente exhibiera de verdad algún patrón común de causalidad. Es decir que Ryle puede afirmar de manera plausible que ningún informe en términos causales podría captar la clase de acciones inteligentes excepto *per accidens*. En apoyo de esa posición —en favor de la cual hay tanto que decir a pesar del embeleso actual por las teorías causales— Ryle puede hacer afirmaciones del tipo de las que Fodor menosprecia ("no es la actividad mental lo que vuelve ingeniosas las actuaciones de los payasos, puesto que lo que las vuelve ingeniosas son hechos tales como que ocurrieron donde los niños podían verlas") sin cometer el error de suponer que las respuestas causales y las conceptuales son incompatibles.[1]

El conductismo lógico de Ryle está viciado por un infundado prejuicio anticientífico, pero no tiene por qué haber sido así. Obsérvese que la intro-

[1] Este párrafo corrige una mala interpretación tanto de la posición de Fodor como de la de Ryle en mi nota crítica del libro de Fodor en *Mind* (1977 - reimpresa en *Brainstorms* págs. 90-108).

ducción al concepto de valencia en química fue una muestra de *conductismo químico lógico:* tener un valencia *n* era "por definición" estar dispuesto a comportarse de tales y cuales formas en tales y cuales condiciones, *como quiera que* esa disposición sea explicada por la física algún día. En este caso determinado, la relación entre la teoría química y la teoría física está actualmente bien trazada y comprendida —aún cuando en medio de los estertores ideológicos la gente a veces describa mal esa relación— y la explicación dada por la física acerca de esas propiedades combinatorias de disposición es un ejemplo primordial de la clase de éxitos científicos que inspiran las doctrinas reduccionistas. Se ha demostrado, que en algún sentido, la química se sometió a la física, y es evidentemente Algo Bueno, es el tipo de cosa de la que tendríamos que tratar de conseguir más.

Semejante progreso insta a la posibilidad de una evolución paralela en psicología. Primero contestaremos la pregunta: "¿Qué tienen en común todos los creyentes en que *p*...?" de la primera manera, la manera "conceptual" y luego veremos si podemos continuar hasta "reducir" la teoría que surge en nuestra primera respuesta a otra cosa: la neurofisiología es la más probable. Muchos teóricos parecen dar por sentado que *parte* de esa reducción es tanto posible como deseable, y tal vez hasta inevitable, aun cuando críticos recientes del reduccionismo, como Putnam y Fodor, nos han advertido acerca de los excesos de los credos reduccionistas "clásicos". Actualmente nadie espera manejar la psicología del futuro con el lenguaje de la neurofisiología, menos aun con el del físico, y se han propuesto formas principios para aflojar las "normas" clásicas de la reducción. El punto es, entonces, *¿qué clase* de vínculos teóricos podemos esperar —o deberíamos poder— encontrar uniendo las afirmaciones psicológicas acerca de las creencias, los deseos y demás con las afirmaciones de los neurofisiólogos, los biólogos y otros científicos de la física?

Puesto que los términos "creencia", "deseo" y sus afines son parte del lenguaje común, como "imán", más que términos técnicos como "valencia", debemos remitirnos a la "psicología popular" para ver qué clase de cosas nos están pidiendo que expliquemos. ¿Qué aprendemos acerca de lo que son las creencias cuando aprendemos a usar las palabras "creer" y "creencia"? El primer punto en el que debemos hacer hincapié es que en verdad no aprendemos qué son las creencias cuando aprendemos a usar esas palabras.[2] Desde luego que nadie *nos dice* qué son las creencias, o si alguien lo hace, o si por casualidad especuláramos sobre el tema por cuenta propia, la respuesta a la que llegamos, sabia o tonta, se traslucirá en forma muy débil en nuestros hábitos de pensar sobre lo que la gente cree. Aprendemos a *utilizar* la psicología popular como una técnica social vernácula, una destreza: pero no la aprendemos conscientemente como teoría —no aprendemos ninguna metateoría junto con la teoría— y a este respecto, nuestros conocimientos de la física popular son como nuestro conocimiento de la gramática de nuestro propio idioma. Este hecho no hace que nuestro conocimiento de la

[2] Creo que vale la pena observar que el uso de "creer" que hacen los filósofos como el término del lenguaje corriente y estándar es una distorsión considerable. Rara vez hablamos de lo que la gente *cree*: hablamos de lo que *piensa* y de lo que *sabe*.

psicología popular sea completamente diferente del conocimiento humano de teorías académicas explícitas, sin embargo; probablemente se podría ser un buen químico en ejercicio y no obstante encontrar embarazosamente difícil lograr una definición escolar satisfactoria de un metal o un ión.

No hay textos de introducción a la psicología popular (aunque a *The Concept of Mind*, de Ryle, podría sacársele el jugo para el efecto), si bien filósofos del idioma común han emprendido muchas exploraciones sobre el tema (con intenciones levemente diferentes), y más recientemente lo han hecho los filósofos de la mente que tienen una mentalidad más teórica, y de todo esto se puede reunir una explicación de la psicología popular en parte axiomática y el resto polémica. ¿Qué son las creencias? A grandes rasgos, la psicología popular dice que las *creencias* son estados de la gente portadores de información, que surgen de las percepciones y que, junto con los *deseos* adecuadamente afines, llevan a la *acción* inteligente. Hasta aquí el tema es relativamente indiscutible, pero ¿dice también la psicología popular que los animales no humanos tienen creencias? Si así fuera, ¿cuál es el papel del lenguaje en la creencia? ¿Están las creencias constituidas por partes? Si así fuera, ¿qué son las partes? ¿Ideas? ¿Conceptos? ¿Palabras? ¿Cuadros? ¿Son las creencias como los actos de habla o como mapas u oraciones de enseñanza? ¿Está implícito en la psicología popular que las creencias forman parte de las relaciones causales o no? ¿Cómo intervienen las decisiones y las intenciones entre los complejos deseo-creencia y las acciones? ¿Son sujeto de introspección las creencias? y si lo son, ¿qué autoridad tienen las declaraciones del creyente?

Todas estas preguntas merecen respuesta, pero es necesario tener en cuenta que hay diferentes razones para estar interesado en los detalles de la psicología popular. Una razón es que existen como fenómeno, como una religión o un idioma o una moda en el vestir, a ser estudiados con las técnicas y las actitudes de la antropología. Puede ser un mito, pero es un mito en el que vivimos, de manera que es un fenómeno "importante" de la naturaleza. Una razón diferente es que parece ser una teoría *verdadera*, de manera general, y de ahí que sea candidata —como la física popular de los imanes y no como la ciencia popular de la astrología— a ser incorporada a la ciencia. Estas razones diferentes generan investigaciones diferentes pero superpuestas. El problema antropológico debería incluir en su explicación de la psicología popular todo lo que el pueblo incluye realmente en su teoría, por más desencaminado e injustificado que pueda ser parte de ella. (Cuando el antropólogo señala parte del catálogo de la teoría popular como falso, puede hablar de un *conocimiento o ideología falsa*, pero el papel de esa teoría falsa *qua* el fenómeno antropológico no disminuye por eso.) Por otra parte, la indagación proto-científica, como tentativa de preparar a la teoría popular para su subsecuente incorporación al resto de la ciencia o para su reducción debería ser crítica y debería eliminar todo lo que es falso o mal fundado, por bien arraigado que esté en la doctrina popular. (Thales creía que las piedras imán tenían alma. Así nos lo dijeron. Aun si la mayor parte de la gente estuviera de acuerdo, esto sería algo que habría que eliminar de la física popular de los imanes antes de la "reducción"). Una manera de distinguir lo bueno de

lo malo, lo esencial de lo injustificado en la teoría popular, es ver qué es lo que hay que incluir en la teoría para explicar cualquier éxito predictivo o explicativo que parezca tener en el uso ordinario. De este modo podemos criticar a medida que analizamos, y hasta tenemos abierta la posibilidad de descartar la psicología popular al final si resulta ser una mala teoría y con ella las entidades presumiblemente teóricas nombradas en ella. Si descartamos la psicología popular como teoría, tendríamos que reemplazarla con otra que, aunque hiciera daños a muchas intuiciones comunes, explicaría el poder predictivo de la habilidad popular residual.

Usamos la psicología popular continuamente para explicar y predecir la conducta mutua; nos atribuimos creencias y deseos los unos a los otros con absoluta seguridad —de manera totalmente inconsciente— y pasamos una parte importante de las horas en las que estamos despiertos ideando el mundo —sin excluirnos— en estos términos. La psicología popular es casi una parte tan penetrante de nuestra segunda naturaleza como lo es nuestra física popular de los objetos de tamaño mediano. ¿Cuán buena es la psicología popular? Si nos concentramos en sus debilidades observaremos que a menudo no podemos encontrarle sentido a determinados aspectos de la conducta humana (incluida la nuestra) en términos de creencia y deseo, ni siquiera retrospectivamente; con frecuencia no podemos predecir de manera exacta o fiable lo que alguien va a hacer o cuándo; frecuentemente no encontramos recursos dentro de la teoría para resolver desacuerdos acerca de determinadas atribuciones de creencia o deseo. Si nos concentramos en sus puntos fuertes descubrimos primero que hay grandes zonas en las que es extraordinariamente fiable en su poder predictivo. Cada vez que nos aventuramos a salir a una autopista, por ejemplo, arriesgamos nuestras vidas a la fiabilidad de nuestras expectativas generales acerca de las creencias perceptivas, los deseos normales y las decisiones de los otros conductores. En segundo término, descubrimos que es una teoría de gran eficiencia y poder generador. Por ejemplo, cuando miramos una película que tiene un argumento muy original y nada estereotipado, vemos al héroe sonreírle al villano y todos llegamos rápidamente y sin esfuerzo al mismo diagnóstico teórico complejo: "¡Ajá!", inferimos (tal vez no conscientemente), "él quiere que ella crea que él no sabe que ella se propone estafar a su hermano". En tercer lugar, descubrimos que hasta los niños pequeños adquieren habilidad con la teoría en un momento en el que tienen una experiencia muy limitada de la actividad humana de la cual inferir una teoría. En cuarto término, descubrimos que todos usamos la psicología popular sin saber casi nada acerca de lo que realmente ocurre dentro del cráneo de la gente. Nos dicen "Usa la cabeza" y sabemos que algunos son más inteligentes que otros, pero nuestra capacidad para utilizar la psicología popular no le afecta para nada la ignorancia acerca de los procesos generales, ni siquiera la mucha información errónea acerca de ellos.

Como lo han observado muchos filósofos, una característica de la psicología popular que la separa tanto de la física popular como de las ciencias físicas académicas es que las explicaciones de acciones en las que se mencionan creencias y deseos, no sólo describen normalmente la procedencia de las acciones, sino que al mismo tiempo las defienden como razonables bajo la circunstancias. Son explicaciones justificativas que hacen una alusión que no se

puede suprimir a la racionalidad del agente. Primordialmente por esta razón, pero también debido al patrón de puntos fuertes y debilidades recién descritas, sugiero que la psicología popular se puede considerar mejor como un cálculo racionalista de interpretación y predicción: un método instrumental de interpretación idealizador y abstracto que ha evolucionado porque funciona y que funciona porque nosotros hemos evolucionado. Nos acercamos los unos a los otros como ***sistemas intencionales*** (Dennett, 1971), es decir como entidades cuya conducta se puede predecir por el método de atribución de creencias, deseos y agudeza racional según los siguientes principios rudimentarios pero eficientes:

1) Las creencias de un sistema son aquellas que ***debería tener***, dadas su capacidad perceptiva, sus necesidades epistemológicas y su biografía. De este modo y en general, sus creencias son tanto reales como inherentes a su vida, y cuando se le atribuyen creencias falsas, hay que contar historias especiales para explicar cómo el error fue el resultado de la presencia de características del entorno que son engañosas en relación con las capacidades perceptivas del sistema.

2) Los deseos de un sistema son aquellos que ***debería tener***, dadas sus necesidades biológicas y los medios más factibles para satisfacerlos. Así es como los sistemas intencionales desean la supervivencia y la procreación y por lo tanto desean comida, seguridad, salud, sexo, riqueza, poder, influencia y así sucesivamente, y también cualquier otra medida que tienda (a su juicio, dadas sus creencias) a favorecer estos fines de manera adecuada. Una vez más se pueden atribuir deseos "anormales" si se pueden contar historias especiales.

3) La conducta de un sistema consistirá en aquellos actos que ***sería racional*** que un agente con esas creencias y deseos ejecutara.

En (1) y (2) "debería tener" significa "tendría si estuviera *idealmente* protegido en su nicho ambiental". De este modo *reconocerá como tales* (es decir, creerá peligrosos) todos los peligros y vicisitudes de su entorno y ***deseará*** todos los beneficios relativos a sus necesidades, por supuesto. Cuando un hecho de su alrededor es especialmente inherente a sus proyectos actuales (que serán los proyectos que un ser así debería tener para progresar en este mundo), *conocerá* ese hecho y procederá en consecuencia. Y así sucesivamente. Esto nos da la noción de un operador o agente epistemológico conativo ideal, reducido a un conjunto de necesidades para la supervivencia y la procreación y al entorno en el que evolucionaron sus antepasados y al cual está adaptado. Pero esta noción es todavía demasiado imperfecta y exagerada. Por ejemplo, un ser puede llegar a tener una necesidad epistemológica que su aparato perceptivo no puede satisfacer (de repente toda la comida verde es venenosa pero es una lástima que él sea daltónico); de ahí la relatividad de las capacidades perceptivas. Más aun, puede haber tenido o no la oportunidad de aprender algo por medio de la experiencia, de manera que sus creencias también están en relación con su biografía de este modo: habrá aprendido lo que debería haber aprendido, viceversa, aquello de lo que había tenido pruebas de forma compatible con su aparato cognitivo, siempre que la evidencia fuera "inherente" a su proyecto en ese momento.

Pero esto es todavía demasiado imperfecto, puesto que la evolución no nos brinda el mejor de todos los mundos posibles sino únicamente un disposi-

tivo provisional, de manera que debamos buscar atajos del diseño que en circunstancias específicamente anormales producen creencias perceptivas falsas, etcétera. (No somos inmunes a las ilusiones, como seríamos si nuestros sistemas perceptivos fueran *perfectos*). Para compensar los atajos en el diseño también deberíamos esperar dividendos de éste: circunstancias en las cuales la manera "barata" que tiene la naturaleza para diseñar un sistema cognitivo tiene el beneficio adicional de dar resultados buenos y fiables aún fuera del entorno en el cual evolucionó el sistema. Nuestros ojos están bien adaptados para darnos creencias verdaderas tanto en Marte como en la Tierra puesto que la solución barata para nuestros ojos de evolución terráquea resulta ser una solución más general (véase Sober, 1981).

Propongo que podemos continuar con ese modo de pensar recién ilustrado *hasta el fin* —no sólo respecto del diseño del ojo, sino también con respecto al diseño de la deliberación, el de la creencia y el de tramar estrategias. Al utilizar este conjunto de presunciones optimistas (la naturaleza nos ha formado para hacer las cosas bien; busque sistemas que crean en la verdad y amen el bien), no les imputamos ningún poder oculto a las necesidades epistémicas, las capacidades perceptivas y la biografía, sino sólo los poderes que el sentido común atribuye a la evolución y el aprendizaje.

En pocas palabras, nos tratamos los unos a los otros como si fuéramos agentes racionales, y este mito —puesto que con seguridad no somos tan racionales— funciona muy bien porque somos *bastante* racionales. Esta única presunción, en combinación con verdades caseras acerca de nuestras necesidades, capacidades y circunstancias típicas, genera tanto una interpretación intencional de nosotros como creyentes y deseadores, como predicciones reales de conducta en gran cantidad. Estoy afirmando, entonces, que la psicología popular se puede considerar mejor como una especie de conductismo lógico: *lo que significa* decir que si alguien cree que p, es así, es que esa persona está dispuesta a conducirse de cierto modo en ciertas condiciones. ¿De qué modo en qué condiciones? En las condiciones en las que sería racional conducirse, dadas las otras creencias y deseos de esa persona. La respuesta parece estar en peligro de ser circular, pero piénselo: una explicación de lo que es para un elemento tener una valencia determinada hará del mismo modo una referencia imposible de eliminar a las valencias de otros elementos. Lo que se nos brinda al hablar de valencias es todo un sistema de atribuciones entrelazadas, que se salva de la vacuidad produciendo predicciones independientemente comprobables.

Acabo de describir a grandes rasgos un método para predecir y explicar la conducta de la gente y otras criaturas inteligentes. Permítame destacar dos preguntas acerca de ella: ¿es algo que podríamos hacer y algo que en realidad hacemos? Creo que la respuesta a la primera es claramente sí, lo que no quiere decir que el método siempre produzca buenos resultados. Tanto como eso se puede asegurar mediante la reflexión y la experimentación sobre el pensamiento. Más aun, se puede reconocer que el método nos es familiar. Aunque no es habitual que usemos el método conscientemente si lo usamos así en aquellas ocasiones en las que la conducta de alguien nos deja perplejos, y entonces a menudo da resultados satisfactorios. Más aun, la facilidad y na-

turalidad con que recurrimos a esta forma consciente y deliberada de resolver problemas brinda cierto apoyo a la afirmación de que lo que hacemos en esas ocasiones no es cambiar de método sino simplemente volvernos conscientes y explícitos acerca de lo que comúnmente logramos tácita o inconscientemente.

Creo que ninguna otra imagen de la psicología popular puede explicar el hecho de que nos vaya tan bien prediciendo las conductas mutuas con pruebas tan pobres y periféricas; tratarnos los unos a los otros como sistemas intencionales funciona (hasta donde lo hace) porque estamos realmente bien diseñados por la evolución y por ello nos *aproximamos* a la versión ideal de nosotros mismos explotados para producir predicciones. Pero la evolución no sólo garantiza que siempre haremos lo que es racional; también garantiza que no lo haremos. Si estamos diseñados por la evolución, entonces ciertamente no somos nada más que un conjunto de mañas, unidas como pedazos por una naturaleza *satisfaciente* —un término de Herbert Simon (1957)— y no mejores de lo que nuestros antepasados tenían que ser para arreglárselas. Más aun, las exigencias de la naturaleza y las de un derrotero lógico no son las mismas. A veces —hasta *normalmente* en ciertas circunstancias, resulta provechoso sacar conclusiones precipitadas o con rapidez (y hasta olvidar que se ha actuado así), de manera que por medio de la mayor parte de las medidas de la racionalidad (consistencia lógica, abstenerse de la deducción carente de validez) ha habido probablemente alguna presión evolucionista positiva a favor de los métodos *irracionales.*[3]

¿Cuán racionales somos? Las investigaciones recientes en psicología social y cognitiva (por ejemplo, Tversky y Kahneman, 1974; Nisbett y Ross 1978) sugieren que somos sólo mínimamente racionales, pasmosamente listos para sacar conclusiones precipitadas o para ser desviados por rasgos lógicamente irrelevantes de las situaciones, pero esta imagen resentida es una ilusión engendrada por el hecho de que estos psicólogos están tratando deliberadamente de producir situaciones que provoquen respuestas irracionales —induciendo a la patología de un sistema tensándolo— y lo logran, pues son buenos psicólogos. Nadie contrataría a un psicólogo para demostrar que la gente elegiría las vacaciones pagadas y no una semana en la cárcel si se le ofreciera una buena información sobre la opción. Al menos no en los mejores departamentos de psicología. Un repaso de las dificultades encontradas por las investigaciones en la inteligencia artificial engendra una impresión más optimista de nuestra racionalidad. Hasta los más sofisticados programas de *IA* caen ciegamente en malas interpretaciones y malentendidos que hasta los niños pequeños eluden confiadamente sin pensarlo dos veces (véase por ejemplo, Schank, 1976; Schank y Abelson, 1977). Desde este lugar privilegiado de observación parecemos maravillosamente racionales.

[3] Mientras que, en general, las verdaderas creencias tienen que ser más útiles que las falsas (y de ahí que un sistema tendría que tener creencias verdaderas), en circunstancias especiales puede ser mejor tener algunas creencias falsas. Por ejemplo, sería mejor para la bestia *B* tener algunas creencias falsas acerca de a quienes puede vencer y *X* no. Clasificando los antagonismos probables de *B* de feroces hasta dominables, ciertamente que queremos que *B* crea que no puede vencer a todas las feroces y a todas las evidentemente poco resistentes, pero es mejor (porque "cuesta menos" en trabajos de discriminación y protege contra las perturbaciones fortuitas tales como los

Por más racionales que seamos, es el mito de nuestra representación racional el que estructura y organiza nuestras atribuciones de creencia y deseo y el que regula nuestras propias delibraciones e investigaciones. Aspiramos a la racionalidad, y sin el mito de nuestra racionalidad los conceptos de creencia y deseo estarían desarraigados. Por lo tanto, la psicología popular está *idealizada* en cuanto a que produce sus predicciones y explicaciones calculando en un sistema normativo; predice lo que vamos a creer, desear y hacer, determinando lo que deberíamos creer, desear y hacer.[4]

La psicología popular es *abstracta* en cuanto a que las creencias y deseos que atribuye no son —o no necesitan ser— supuestos como estados destacados interpuestos de un sistema causante de conducta interno. (Se ampliará el tema más adelante.) El papel del concepto de creencia es como el papel del concepto de un centro de gravedad y los cálculos que producen las predicciones son más como los cálculos que se hacen con un paralelogramo de fuerzas que los que se hacen con una copia en papel carbón de las palancas y los engranajes internos.

La psicología popular es por tanto *instrumental* de una manera que los realistas más ardientes deberían permitir: la gente tiene verdaderamente creencias y deseos en mi versión de la psicología popular, del mismo modo que tienen realmente centros de gravedad y la Tierra tiene un ecuador.[5] Reichenbach distinguía entre dos clases de referentes para los términos teóricos: *illata* —entidades teóricas postuladas— y *abstracta*, entidades limitadas por los cálculos o conceptos lógicos.[6] Las creencias y deseos de la psicología popular (pero no todos los hechos y estados mentales) son *abstracta*.

Esta perspectiva de la psicología popular surge más claramente cuando se la contrasta con un panorama diametralmente opuesto cuyos dogmas han sido sostenidos por algún filósofo, y al menos la mayoría de los cuales han sido adoptados por Fodor:

---

días malos y los golpes de suerte) para *B* extender el "No puedo vencer a X" hasta abarcar algunas bestias a las que en realidad puede vencer *Errar del lado de la prudencia* es una buena estrategia bien reconocida, y se puede esperar que la naturaleza lo haya valorado así en las ocasiones en que apareció. Una estrategia alternativa sería en este caso seguir la regla: evitar conflictos con los casos oscuros. Pero habría que "pagar más" para implementar esa estrategia que para implementar la estrategia diseñada para producir ciertas falsas creencias y confiar en ellas. (Acerca de las creencias falsas, véase también el capítulo 2.)

[4] Prueba sus predicciones de dos maneras: prueba las predicciones de acción directamente viendo lo que el agente hace; las predicciones de creencia y deseo se prueban indirectamente empleando las predicciones atribuidas en predicciones ulteriores de acción eventual. Como siempre, la tesis duhemiana se sostiene: las atribuciones de creencia y deseo están poco determinadas por los datos disponibles.

[5] La "Theoretical Explanation" de Michael Fridman (1981) proporciona un excelente análisis del papel del pensamiento instrumental en la ciencia realista. Scheffler (1963) hace una distinción útil entre *instrumentalismo* y *ficcionalismo*. En sus términos estoy caracterizando a la psicología popular como instrumental, no ficcional.

[6] Nuestras observaciones de las cosas concretas confieren cierta probabilidad a la existencia de *illata*, nada más. Segundo, hay inferencias de las *abstracta*. Estas inferencias son... equivalencias, no inferencias de probabilidad. En consecuencia, la existencia de abstracta se puede reducir a la existencia de concreta. Por lo tanto, no existe ningún problema acerca de sus existencia objetiva; su status depende de una convención." (Reichenbach, 1938, págs. 211-12.)

> Las creencias y los deseos, tal como los dolores, los pensamientos, las sensaciones y otros episodios, son considerados como eventos o estados internos reales e *interpuestos* en la intención causal, incluidos en leyes protectoras de índole causal. La psicología popular no es un cálculo idealizado y racional sino una teoría naturalista, empírica y descriptiva, que imputa regularidades causales descubiertas por la inducción extensiva sobre la experiencia. Suponer que dos personas comparten una creencia es suponer que están finalmente en algún estado interno estructuralmente similar, por ejemplo, que tienen las mismas palabras del lenguaje mentalista escritas en los lugares funcionalmente pertinentes de sus cerebros.

Quiero desviar este choque frontal de los análisis tomando dos medidas. Primero, estoy preparado para conceder cierto grado a las afirmaciones hechas por la oposición. Por supuesto que no estamos todos sentados en la oscuridad en nuestros estudios como leibnicianos locos inventando de manera racional predicciones conductistas sacadas de conceptos puros e idealizados de nuestros vecinos, ni derivamos nuestra celeridad para atribuir deseos de una generación cuidadosa de ellos a partir de la última meta de la supervivencia. Podemos observar que algunas personas parecen desear cigarrillos, o dolor, o notoriedad (esto lo observamos oyéndolos contárnoslo, mirando lo que eligen, etc.) y sin ninguna convicción de que esta gente, dadas sus circunstancias debería tener estos deseos, se los atribuimos igual. De este modo la generación racionalista de atribuciones aumenta y hasta se corrige en ocasiones por medio de generalizaciones empíricas acerca de la creencia y el deseo, que guían nuestras atribuciones y que se aprenden más o menos en forma inductiva. Por ejemplo, los niños pequeños creen en Papá Noel, la gente se siente inclinada a creer en la que le sea más útil de dos interpretaciones de un hecho en el que están involucrados (a menos que estén deprimidos), y a la gente se le puede hacer desear cosas que no necesitan haciéndole creer que a la gente encantadora le gustan esas cosas. Y así sucesivamente, con la abundancia conocida. Este folclore no consiste en *leyes* —ni siquiera leyes de probabilidad— sino que parte de él está siendo convertido en una ciencia de cierto tipo, por ejemplo teorías de "cognición caliente" y disonancia cognitiva. Concedo la existencia de toda esta generalización naturalista y su papel en los cálculos normales de los psicólogos populares, es decir, todos nosotros. La gente confía en su propio grupo parroquial de vecinos cuando formula interpretaciones intencionales. Esa es la razón por la cual se tiene tanta dificultad en entender a los extranjeros: su conducta, por no decir nada de sus idiomas. Imputan más de sus propias creencias y deseos y los de sus vecinos, que los que imputarían si siguieran servilmente mis principios de atribución. Por cierto, que éste es un atajo perfectamente razonable que la gente toma, aun cuando con frecuencia conduce a malos resultados. En este tema, como en la mayoría de ellos, somos satisfactores, no optimizadores, cuando se trata de reunir información y construir teorías. Sin embargo yo insistiría en que todo este saber popular obtenido empíricamente está colocado sobre un armazón generador y normativo fundamental que tiene las características que he descrito.

Mi segunda medida, lejos del conflicto que he señalado, es recordar que el punto no es qué es realmente la psicología popular tal como se la en-

cuentra en este campo, sino lo que es en el mejor de los casos; qué es lo que merece ser tomado en serio e incorporado a la ciencia. No se trata especialmente de argumentar contra mí que la psicología popular está *en realidad* comprometida con las creencias y deseos como *illata* perceptible, de interacción causal; lo que se debe señalar es que debería serlo. Trataré esta última afirmación a su debido tiempo. *Podría* conceder la primera afirmación sin enturbiar a mi proyecto total, pero no lo hago, puesto que me parece una prueba bastante fuerte el que nuestra noción común de creencia no tiene casi nada de lo concreto en ella. Jacques mata a su tío de un tiro en la Plaza Trafalgar y es arrestado en el acto por Sherlock; Tom lo lee en el *Guardian* y Boris se entera de ello por el *Pravda*. Ahora bien, Jacques, Sherlock, Tom y Boris han tenido experiencias notablemente distintas —sin mencionar sus biografías anteriores y proyectos futuros—pero hay algo que comparten: todos creen que un francés ha cometido un asesinato en la Plaza Trafalgar. Todos no *dijeron* esto, ni siquiera se lo dijeron "a ellos mismos". Podemos suponer que esa *proposición* "no se le ocurrió" a ninguno de ellos, y aunque se les hubiera ocurrido, habría tenido un significado completamente distinto para Jacques, Sherlock, Tom y Boris. Sin embargo, todos creen que un francés cometió un asesinato en la Plaza Trafalgar.Esta es una propiedad compartida que es visible, por así decirlo, sólo desde un punto de vista muy limitado: el punto de vista de la psicología popular. Los psicólogos populares comunes no tienen ninguna dificultad en imputarle a la gente esa vulgaridad tan útil como evasiva. Si insisten en que, al hacerlo están postulando un objeto estructurado de manera similar en cada cabeza, ésta es una concreción injustificada y mal colocada, un desliz lamentable de la ideología.

Pero de cualquier manera no hay duda de que la psicología popular es una mezcolanza, como todos los productos populares, y no hay ningún motivo para no admitir, al fin, que es mucho más compleja, abigarrada (a riesgo de ser incoherente) de lo que mi esbozo la ha hecho parecer. La noción *corriente* de creencia coloca sin duda a las creencias en alguna parte de la mitad del camino entre ser *illata* o *abstracta*. Esto me sugiere que el concepto de creencia que se encuentra en la comprensión común, es decir en la psicología popular, es poco atractivo como concepto científico. Me recuerda al extraño precursor del atomismo de Anaxágoras: la teoría de las semillas. Hay una parte de todo en todo, se dice que afirmó. Todos los objetos consisten en una infinidad de semillas, de todas las variedades posibles. ¿Cómo se hace el pan con harina, levadura y agua? La harina contiene semillas de pan en abundancia (pero predominan las semillas de harina; eso es lo que la convierte en harina), y lo mismo pasa con la levadura y el agua, y cuando estos ingredientes se mezclan, las semillas de pan forman una mayoría nueva, de manera de que lo que se obtiene es pan. El pan nutre por tener semillas de carne, sangre y hueso además de su mayoría de semillas de pan. Estas semillas no son buenas entidades teóricas, puesto que, al ser una especie de cruce bastardo entre las propiedades y las partes apropiadas tienen propensión a generar regresiones perversas, y sus condiciones de identidad son problemáticas, por decir sólo algo.

Las creencias son algo así. No parece existir ninguna manera cómoda de evitar la afirmación de que tenemos infinidad de creencias, y la intuición co-

mún no nos da una respuesta constante a acertijos tales como la creencia de que si 3 es mayor que 2 no es otra que la creencia de que 2 es menor que 3. La respuesta evidente al desafío de infinidad de creencias con condiciones de identidad evasivas es suponer que estas creencias no están todas "almacenadas por separado"; muchas —en realidad la mayor *parte* si estamos hablando verdaderamente de lo infinito— están almacenadas *implícitamente* en virtud del almacenaje *explícito* de unas pocas (o unos pocos millones): las *creencias núcleo* (véase Dennett 1975; también Fodor 1975 y Field 1978). Las creencias núcleo "se almacenarán por separado", y parecen *illiata* promisorias en contraste con las creencias virtuales o implícitas que se parecen a las *abstracta* paradigmáticas. Pero aunque ésta resultara ser la forma en que están organizados nuestros cerebros, sospecho que las cosas serán más complicadas que esto: no hay ninguna razón para suponer que los *elementos* núcleo, los signos de representación concretos, destacados, almacenados por separado (y debe de haber algunos de esos elementos en cualquier sistema complejo de procesamiento de información), representarán explícitamente (o *serán*) un subconjunto de nuestras *creencias*, siquiera. Es decir que si uno se sentara y escribiera una lista de unas mil creencias paradigmáticas propias, *todcs* ellas podrían resultar ser virtuales y sólo implícitamente almacenadas o representadas, y lo que estuviera explícitamente almacenado sería información completamente conocida (por ejemplo acerca de direcciones memorizadas, procedimientos para resolver problemas. o reconocimiento, etc.). Sería insensato prejuzgar este tema empírico insistiendo en que nuestras representaciones núcleo de información (cualesquiera que resulten ser) son creencias *par excellence*, puesto que cuando los hechos están, nuestras intuiciones pueden apoyar en cambio el punto de vista contrario: las auto-atribuciones de creencias menos polémicas pueden distinguir creencias que desde la posición privilegiada de la teoría cognitiva desarrollada son invariablemente virtuales.[7]

En ese caso, ¿qué podríamos decir acerca de los papeles causales que les asignamos comúnmente a las creencias (por ej., "La creencia que ella tenía de que John conocía su secreto la hacía ruborizarse")? Podríamos decir que cualesquiera que fueran los elementos núcleo en virtud de los cuales ella creía virtualmente que John conocía su secreto, ellos, los elementos núcleo desempeñaban un papel causal directo (en cierto modo) en provocar el enrojecimiento como respuesta. Sería prudente, como lo prueba este ejemplo, no manosear nuestro catálogo *común* de creencias (aunque todas resultaran ser virtuales), pues éstas son regularidades, predecibles, fáciles de entender y manipulables de los fenómenos psicológicos a pesar de su aparente neutralidad con respecto a la distinción explícita/implícita (o central/virtual). Lo que Jacques, Sherlock, Tom y Boris tienen en común es probablemente sólo una creencia virtual "derivada" de muy diferentes almacenes explícitos de información en cada uno de ellos, pero virtuales o no, es el hecho de compartir *esta* creencia lo que explicaría (o nos permitiría predecir) en determinadas circunstancias imaginarias, que todos ejecutarían la misma acción cuando se

[7] Véase Field, 1978, pág. 5, nº 12 acerca de "concesiones menores" a esos tratamientos instrumentales de la creencia.

les diera la misma información nueva. (Y ahora por un millón de dólares, Tom [o Jacques, Sherlock, Boris], conteste correctamente nuestra pregunta por el lote acumulado: ¿alguna vez un ciudadano francés cometió un crimen importante en Londres?".)

Al mismo tiempo deseamos aferrarnos a la idea igualmente común de que las creencias pueden causar no sólo acciones, sino también rubores, deslices verbales, ataques cardíacos, y demás. Gran parte de la discusión acerca de si las explicaciones intencionales son o no explicaciones causales se pueden mencionar como desempeñando el papel causal, mientras que la creencia permanece virtual. "Si Tom no hubiera creído que *p* y hubiera querido que *q*, no habría hecho *A*." ¿Es ésta una explicación causal? Es equivalente a esto: "Tom estaba en alguno de un número indefinido de estados de tipo *B* estructuralmente diferentes que tienen en común sólo que cada uno de ellos autoriza la atribución de creencia de que *p* y el deseo de que *q* en virtud de sus relaciones normales con muchos otros estados de Tom, y este estado, cualquiera que haya sido, fue causalmente suficiente, dadas las "condiciones de fondo" por supuesto, de iniciar la intención de ejecutar *A* y, por consiguiente *A* se ejecutó y si *él* no hubiera estado en uno de esos infinitos estados de tipo *B*, no lo hubiera hecho. Se la puede llamar explicación causal porque habla de causas, pero es seguramente todo lo poco específica e inútil que una explicación causal puede llegar a ser. Se compromete a que haya una u otra explicación que caiga dentro de un área muy amplia (se considera a la interpretación intencional como superviniente en el estado corporal de Tom), pero su verdadera utilidad y valor informativo en la predicción real se halla, no de modo sorprendente, en su afirmación de que Tom, como sea que su cuerpo esté estructurado en el momento, tiene un conjunto especial de estas evasivas propiedades, creencias y deseos intencionales.

La idea común de creencia se bifurca en dos direcciones. Si queremos tener buenas entidades teíricas, buenas *illata*, o buenas construcciones lógicas, buenas *abstracta*, tendremos que deshacernos de parte de la carga ordinaria de los conceptos de creencia y deseo. Es así como propongo un divorcio. Puesto que parece que tenemos ambas ideas casadas con la psicología popular, debemos separarlas y crear dos teorías nuevas: una estrictamente abstracta, idealizadora, holística, instrumental —la teoría pura del sistema intencional— y la otra una ciencia concreta, microteórica de la verdadera comprensión de esos sistemas intencionales, lo que llamaré psicología cognitiva subpersonal. Explorando sus diferencias e interrelaciones, deberíamos poder decir si hay "reducciones" plausibles en perspectiva.

La teoría del sistema intencional como una teoría de competencia

La primera teoría nueva, la teoría del sistema intencional, se visualiza como una pariente cercana de, y superpuesta con, disciplinas ya existentes como la teoría de la decisión y la teoría del juego, que son igualmente abstractas, normativas y expresadas en el lenguaje intencional. Toma prestados los términos comunes "creencia" y "deseo" pero les da un significado técnico dentro de la teoría. Es una especie de conductismo holístico lógico porque

trata la predicción y la explicación desde los perfiles creencia-deseo de las acciones de sistemas totales (ya sea solos en entornos o en interacción con otros sistemas intencionales), pero trata las realizaciones individuales de los sistemas como cajas negras. El *sujeto* de todas las atribuciones intencionales es la totalidad del sistema (la persona, el animal o hasta la corporación o nación [véase Dennett, 1976]) más que cualquiera de sus partes, y las creencias y deseos individuales no son atribuibles aisladamente, independientemente de otras atribuciones de creencia y deseo. El último punto distingue claramente la teoría del sistema intencional del conductismo lógico de Ryle, que trató de explicar las creencias individuales (y otros estados mentales) como disposiciones particulares individuales hacia la conducta exterior.

La teoría se ocupa de la "producción" de nuevas creencias y deseos a partir de los viejos por medio de una interacción entre viejas creencias y deseos, características del entorno y las acciones del sistema; y esto crea la ilusión de que la teoría contiene descripciones naturalistas del procesamiento interno en los sistemas de los que se ocupa la teoría, cuando en realidad el procesamiento está totalmente en la manipulación de la teoría y consiste en poner al día la caracterización intencional de todo el sistema de acuerdo con las normas de la atribución. Una ilusión análoga de proceso le podría ocurrir a un estudiante ingenuo, quien, al ser confrontado con un paralelogramo de fuerzas supusiera que dibujaba un encadenamiento mecánico de varillas y pivotes de algún tipo en lugar de ser sencillamente una manera gráfica de representar y diseñar el efecto de varias fuerzas actuando simultáneamente.

Richard Jeffrey (1970), al desarrollar sus conceptos de cinemática de probabilidades, ha atraído de forma útil la atención provechosamente hacia una analogía con la diferenciación en física entre cinemática y dinámica. En cinemática,

> se habla de la propagación de los movimientos a través de un sistema en términos de restricciones tales como la rigidez y la forma de eslabonamiento. Es la física de la posición y el tiempo, en cuyos términos de los cuales se puede hablar de velocidad y aceleración, pero no acerca de fuerza y masa. Cuando se habla de fuerzas —***causas*** de aceleraciones— se está en el reino de la dinámica (pág. 172).

La cinemática proporciona un nivel de abstracción simplificado e idealizado adecuado para muchos fines —por ejemplo para el desarrollo del diseño inicial de una caja de cambios— pero cuando uno debe tratar con detalles más concretos de los sistemas —cuando el diseñador de la caja de cambios tiene que preocuparse por la fricción, la flexión, la eficiencia energética y demás, uno debe pasarse a la dinámica en busca de predicciones más detalladas y fiables, al precio de una complejidad aumentada y de una generalidad disminuida. Del mismo modo, uno puede enfocar el estudio de la creencia (el deseo y así sucesivamente) a un nivel, sumamente abstracto, pasando por alto problemas de realización y exponer simplemente cuáles son las exigencias normativas del diseño de un creyente. Por ejemplo, podemos formular preguntas tales como "¿Cuáles deben ser las capacidades y propensiones epistemológicas de un sistema para que sobreviva en el entorno *A*?" (véase Campbell, 1973, 1977) o "¿Qué tiene que saber ya este sistema para que

pueda aprender *B*?" o "¿Qué intenciones debe tener este sistema para significar algo al decir algo?".

La teoría del sistema intencional se ocupa únicamente de las especificaciones del desempeño de los creyentes mientras permanece en silencio acerca de cómo deben implementarse los sistemas. En realidad esta neutralidad acerca de la implementación es la característica más útil de las caracterizaciones intencionales. Considérese, por ejemplo, el papel de las caracterizaciones intencionales en biología evolutiva. Si hemos de explicar la evolución de capacidades conductistas complejas o de talentos cognitivos por selección natural debemos notar que es la capacidad intencionalmente caracterizada (por ejemplo, la capacidad de adquirir una creencia, un deseo, de ejecutar una acción intencional) la que tiene valor de supervivencia como sea que resulte realizada como resultado de la mutación. Si un insecto particularmente nocivo hace su aparición en un entorno, los pájaros y murciélagos que tienen una ventaja de supervivencia serán aquellos que lleguen a creer que no es bueno comerse este insecto. En vista de las amplias diferencias en la estructura neuronal, los antecedentes genéticos y la capacidad perpetua entre pájaros y murciélagos, es sumamente improbable que este rasgo útil que puedan llegar a compartir, tenga una descripción común en algún nivel más completo o menos abstracto que la teoría del sistema intencional. No se trata sólo de que el predicado intencional sea un predicado proyectable en la teoría evolutiva; puesto que es más general que sus predicados que son su contraparte específica de especies (que caracterizan la mutación feliz sólo en los pájaros o en los murciélagos) es preferible. De manera que desde el punto de vista de la biología evolutiva, no querríamos "reducir" todas las caracterizaciones intencionales aunque supiéramos en casos determinados cuál era la implementación psicológica.

Este nivel de generalidad es esencial si queremos que una teoría tenga algo significativo y defendible que decir acerca de temas tales como la inteligencia en general (como opuesta, digamos, a la simple inteligencia humana o hasta terrestre o natural) o temas tan grandiosos como el significado o la referencia o la representación. Supóngase, para proseguir con un tema filosófico conocido que somos invadidos por marcianos y surge la pregunta: ¿Tienen creencias y deseos? ¿Son tan *parecidos a nosotros*? Según la teoría del sistema intencional, si estos marcianos son lo suficientemente listos como para llegar hasta aquí, luego muy ciertamente tienen creencias y deseos en el sentido técnico patentado en la teoría, no importa cuál sea su estructura interna, ni cómo nuestras intuiciones psicológicas populares se rebelen ante la idea.

Este principio negro de la teoría del sistema intencional ante la estructura interna, parece invitar a réplica; pero tiene que haber *alguna* explicación del *éxito* de la predicción intencional de la conducta de sistemas (por ejemplo Fodor, 1985, pág. 79). No es simplemente magia. No es una mera coincidencia que se puedan generar todas estas *abstracta*, manipularlas por la vía de alguna versión del razonamiento práctico y aparecer con una predicción de acción que tiene muy buena probabilidad de ser cierta. Debe haber alguna manera en la cual los procesos internos del sistema reflejen las complejidades de la interpretación intencional. Si no, su éxito sería un milagro.

Por supuesto, todo esto es completamente cierto e importante. Nada que no tenga una gran cantidad de complejidad estructural y de procesamiento, podría concebiblemente comprender un sistema intencional de algún interés, y la complejidad de la comprensión seguramente se parecerá notablemente a la complejidad de la interpretación instrumental. Del mismo modo, el éxito de la teoría de las valencias en química no es una coincidencia y la gente tenía toda la razón en esperar que las semejanzas microfísicas profundas fueran descubiertas entre los elementos con la misma valencia y que las semejanzas estructurales encontradas, explicaran las semejanzas disposítivas. Pero puesto que la gente y los animales no son como los átomos y las moléculas no sólo por ser los productos de una compleja historia evolutiva, sino también por ser los productos de sus historias individuales de aprendizaje, no hay ninguna razón para suponer que los cryeentes individuales (humanos) de que *p* —como átomos individuales (de carbono) con la valencia 4— regulan sus disposiciones con *exactamente* la misma maquinaria. Demostrar las presiones sobre el diseño y la variación de la implementación, y demostrar cómo especies e individuos particulares en realidad tienen éxito en comprender los sistemas intencionales, es la tarea para la tercera teoría: la psicología cognitiva subpersonal.

La psicología cognitiva subpersonal como teoría de ejecución

La tarea de la psicología cognitiva subpersonal es explicar algo que a primera vista parece completamente misterioso e inexplicable. El cerebro, como la teoría del sistema intencional y la biología evolutiva nos lo demuestran, es una *máquina semántica*; su trabajo es descubrir lo que *significan* sus múltiples entradas de datos, discriminadas según su significación y "proceder en consonancia".[8] Eso es *para lo que sirve* el cerebro. Pero el cerebro, como la fisiología o el simple sentido común nos lo demuestran, no es más que una *máquina sintáctica;* todo lo que puede hacer es discriminar sus entradas de datos por sus características estructurales, temporales y físicas y dejar que la totalidad de sus actividades mecánicas sean gobernadas por estas características "sintácticas" de sus entradas de datos. Eso es todo lo que el cerebro *puede hacer*. Ahora bien, ¿cómo se las arregla el cerebro para obtener semántica de la sintaxis? ¿Cómo podría *cualquier* entidad (cómo podría un genio o un ángel o Dios) obtener la semántica de un sistema de nada más que su sintaxis? No podría. La sintaxis de un sistema no determina su semántica. ¿Entonces, por medio de qué alquimia extrae el cerebro resultados semánticamente fiables de operaciones sintácticamente impulsadas? No se lo puede diseñar para realizar una tarea imposible, pero podría diseñárselo para

[8] De manera más exacta, si bien menos pintoresca, la tarea del cerebro es llegar a producir respuestas internas mediadoras que varían sistemáticamente en concierto con la variación en el significado ambiental real (los significados naturales y no naturales, en el sentido de Grice (1957) de sus causas distales e independientemente de variaciones irrelevantes de significados en sus causas contiguas, y más aun, para responder a sus propias respuestas mediadoras de manera que tienden sistemáticamente a mejorar las perspectivas de la criatura en su entorno si las repuestas mediadoras varían como lo deberían hacer.

aproximarse a la tarea imposible, para *remedar* la conducta del objeto imposible (la máquina semántica) capitalizando las correspondencias fortuitas aproximadas (suficientemente aproximadas) entre las regularidades estructurales —del entorno y de sus propios estados y operaciones internas— y los tipos semánticos.

La idea básica es conocida. Un animal necesita saber cuándo ha satisfecho la meta de encontrar e ingerir comida, pero se conforma con un detector de una fricción en la garganta seguida por el estómago dilatado, una sustitución mecánica provocada por un estado mecánico simple que normalmente co-ocurre con la satisfacción de la meta "real" del animal. No es una fantasía y se puede explotar fácilmente para conseguir que el animal coma cuando no debe o que deje de comer cuando no debe, pero funciona bastante bien en el caso del animal y su entorno normal. O supóngase que estoy verificando transmisiones telegráficas que se me ha pedido que intercepte todas las *amenazas de muerte* (pero sólo las amenazas de muerte en inglés para que sea "fácil"). Yo quisiera fabricar una máquina que me ahorrara la molestia de interpretar semánticamente todos los mensajes enviados, ¿pero cómo se podría hacer esto? No se podría diseñar ninguna máquina que hiciera el trabajo perfectamente, pues eso exigiría definir la categoría semántica *amenaza de muerte en inglés* como una característica tremendamente compleja de ristras de símbolos alfabéticos y no hay absolutamente ninguna razón para suponer que esto se podría hacer de forma justificada. (Si de algún modo por la inspección brutal y enumeración subsiguiente pudiéramos hacer una lista de solamente todas las amenazas de muerte en inglés de, digamos, menos de mil signos alfabéticos, podríamos construir bastante fácilmente un filtro que las detectara, pero estamos buscando un método con principios, proyectable y extensible). Se podría armar un artefacto verdaderamente tosco para discriminar todos los mensajes que contuvieran las ristras de símbolos

...te voy a matar...

o

...tú ...mueres ...a menos que...

o

...(para cualquier disyunción limitada de modelos probables que se encuentran en las amenazas de muerte en inglés).

Este dispositivo tendría cierta utilidad, y posteriores refinamientos podrían seleccionar el material que pasara este primer filtro, y así sucesivamente. Es un comienzo poco prometedor para construir un entendedor de oraciones, pero si se quiere obtener semánticas de la sintaxis (ya sea la sintaxis de los mensajes en un lenguaje natural o la sintaxis de los impulsos aferentes de las neuronas), las variaciones de esta estrategia básica son la única esperanza.[9] Hay que juntar un montón de trucos y esperar que la naturaleza

[9] Se podria pensar que mientras en principio no se puede derivar la semántica de un sistema de nada más que de su sintaxis, en la práctica uno podría hacer un poco de trampa y explotar las características sintácticas que no implican una interpretación semántica pero que la sugieren con fuerza. Por ejemplo, enfrentados con la tarea de descifrar documentos aislados en un idioma enteramente desconocido y extraño, no podría notar que mientras que el símbolo que se parece a un

sea lo suficientemente generosa como para permitir que este dispositivo consiga pasar. Naturalmente algunos trucos son elegantes y recurren a profundos principios de organización, pero al final sólo se puede tener la esperanza de producir (todo lo que la selección natural puede haber producido) sistemas *que parecen* discriminar significados discriminando en realidad cosas (rasgos de tipos sin duda localmente disyuntivos) que co-varían fiablemente con los significados.[10] La evolución ha diseñado nuestros cerebros no sólo para hacer esto sino también para evolucionar y seguir estrategias de automejoramiento en esta actividad durante sus vidas individuales (véase Dennett, 1974b).

Es la tarea de la psicología cognitiva subpersonal proponer y probar modelos de tal actividad —reconocimiento de modelos o generalización de estímulos, aprendizaje de conceptos, expectativas, aprendizaje, conducta dirigida hacia cierto fin, solución de problemas— que no sólo produce un simulacro de sensibilidad genuina del contenido, sino que lo hace de manera demostrablemente semejante a la manera en que lo hace el cerebro de la gente, exhibiendo los mismos poderes y las mismas vulnerabilidades ante la decepción, la sobrecarga y la confusión. Es aquí donde encontraremos nuestras buenas entidades teóricas, nuestras ***illata*** útiles, y mientras algunas de ellas pueden parecerse a las entidades conocidas de la psicología popular —creencias, deseos, juicios, decisiones— muchas no se parecerán (véase, por ej., los estados sub-doxásticos propuestos por Stich, 1978b). La única semejanza que podemos estar seguros de descubrir en las ***illata*** de la psicología cognitiva sub-personal es la intencionalidad de sus rótulos (véase ***Brainstorms***, págs. 23-38). Estarán caracterizados como hechos con contenido que portan informaciones, que señalan esto y ordenan aquello.

Para darles a las ***illata*** estos rótulos, para mantener cualquier interpretación intencional de su funcionamiento, el teórico debe siempre echar una

---

pato no tiene que significar "pato", hay una buena posibilidad de que signifique eso, especialmente si el símbolo que parece un lobo puede estar comiéndose el símbolo que se asemeja a un pato y no a la inversa. Llamemos a esto confiar en *los jeroglíficos* y notemos que la forma que ha tomado en las teorías psicológicas desde Locke hasta la actualidad: podremos distinguir las representaciones mentales (cuál idea es la idea de perro y cuál la de gato) porque la primera se asemejará a un perro y la segunda será parecida a un gato. Esto nos viene muy bien como muleta ya que nosotros los observamos desde afuera, que tratamos de asignar contenido a lo que ocurre en algún cerebro, pero el cerebro no le sirve... ¡porque el cerebro no sabe cómo son los perros! Mejor aun, éste no puede ser el método fundamental del cerebro para extraer clases semánticas de la sintaxis no elaborada, puesto que cualquier cerebro (o parte del cerebro) del cual se pudiera decir —en un sentido amplio— que sabe cómo son los perros, sería un cerebro o parte de un cerebro que ya había resuelto su problema, que ya era (un simulacro de) una máquina semántica. Pero esto todavía es engañoso puesto que en ningún caso los cerebros *asignan* contenido a sus propios hechos del modo en que podrían hacerlo los observadores: el cerebro *fija* el contenido de sus hechos internos en el acto de reaccionar como lo hacen. Hay buenas razones para postular *imágenes mentales* de una clase u otra en las teorías cognitivas (véase "Two Approaches to Mental Images" en ***Brainstorms***, pág. 174-89) pero confiar en los jeroglíficos no es una de ellas, aunque sospecho que tiene una influencia disimulada.

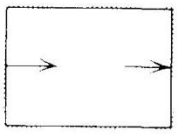
[10] Considero que este punto está íntimamente relacionado con las razones que tiene Davidson para afirmar que no puede haber leyes psicofísicas, pero dudo de que Davidson quiera sacar de él las mismas conclusiones que yo. (Véase Davidson, 1970).

mirada fuera del sistema para ver qué es lo que normalmente produce la configuración que está describiendo, qué efectos tienen normalmente las respuestas del sistema sobre el entorno, y qué beneficio para la totalidad del sistema resulta normalmente de esta actividad. En otras palabras el psicólogo cognitivo no puede pasar por alto el hecho que lo que está estudiando es la comprensión de un sistema intencional so pena de abandonar la interpretación semántica y por lo tanto la psicología. Por otra parte el progreso en la psicología personal subcognitiva borrará los límites entre ella y la teoría del sistema intencional, entretejiéndolas tanto como se entretejieron la química y la física.

La alternativa de ignorar el mundo exterior y sus relaciones con el mecanismo interno (lo que Putnam llamó psicología en un sentido restringido o solipsismo metodológico y que Gunderson ridiculizó como perspectivismo de la caja negra de cristal del mundo) no es en realidad psicología en absoluto, sino simplemente, y en el mejor de los casos, neurofisiología abstracta, pura sintaxis interna sin ninguna esperanza de interpretación semántica. La psicología "reducida" a la naueofisiología de esta manera no sería psicología, puesto que no podría dar una explicación de las regularidades que es la tarea especial de la psicología; la fiabilidad con la cual los organismos "inteligentes" pueden arreglárselas con sus entornos y así prolongar sus vidas. La psicología puede, y debería, trabajar hacia una explicación de los fundamentos fisiológicos de los procesos psicológicos, no eliminando las caracterizaciones psicológicas o intencionales de esos procesos, sino mostrando cómo el cerebro implementa las especificaciones de ejecución intencionalmente caracterizadas de las teorías subpersonales.

Conductismo de la caja negra

Perspectivismo de la caja negra de cristal del mundo



Friedman, al discutir la perplejidad actual de la psicología cognitiva sugiere que el problema

> es de dirección de la reducción. La psicología contemporánea trata de explicar la actividad cognitiva *individual* independientemente de la actividad cognitiva social, y entonces trata de dar una *micro*-reducción de la actividad cognitiva social —es decir, el uso de un lenguaje público— en términos de una teoría anterior de la actividad cognitiva individual. La sugerencia opuesta es que busquemos pri-

mero una teoría de la actividad social y tratemos entonces de dar una *macro*-reducción de la actividad cognitiva individual —la actividad de aplicar conceptos, formular juicios, y demás— en términos de nuestra teoría social anterior (1981, págs. 15-16).

Estoy de acuerdo en general con la idea de la macro-reducción en psicología, excepto en que en la identidad de Friedman del macro-nivel como explícitamente social es sólo una parte de la historia. Las capacidades cognitivas de los animales que no usan un lenguaje (y los Robinson Crusoe, si es que los hay) también deben ser tenidas en cuenta, y no sólo en términos de analogía con las prácticas de nosotros, los que usamos un idioma. El macro-nivel *hasta* el cual deberíamos vincular los micro procesos del cerebro para entenderlos como psicológicos es más ampliamente el nivel de la interacción, desarrollo y evolución organismo-entorno. Ese nivel incluye la interacción social como parte especialmente importante (véase Burge, 1979), pero aun así una parte apropiada.

No hay manera de captar las propiedades semánticas de las cosas (los símbolos de palabras, los diagramas, los impulsos nerviosos, los estados cerebrales) mediante una micro-reducción. Las propiedades semánticas no son sólo de relación sino, podría decirse, de súper-relación, puesto que la relación de un vehículo de contenido determinado, o símbolo, debe portar no es sólo la relación que mantiene con otras cosas semejantes (por ejemplo, otros símbolos, o partes de símbolos, o conjuntos de símbolos o causas de símbolos) sino una relación entre el símbolo y la totalidad de la vida —y de la vida contraobjetiva—[11] del organismo al que "sirve" *y* los requerimientos de ese organismo para la supervivencia *y* su ancestro evolucionista.

Las perspectivas de la reducción

De nuestras tres psicologías —la psicología social, la teoría del sistema intencional y la psicología cognitiva sub-personal—, ¿qué podría reducirse a qué? Por cierto que la micro-reducción de un paso de la psicología popular a la fisiología aludida en los eslogans de los primeros teóricos de la identidad nunca se encontrará y no debería ser pesada por alto, ni siquiera por los partidarios del materialismo y la unidad científica. Una perspectiva digna de explorarse, sin embargo, es la de que la psicología popular (más precisamente la parte de psicología popular de la que vale la pena ocuparse) se reduce a la teoría del sistema intencional. Lo que esto significaría se puede sacar a relucir mejor contrastando esta reducción conceptual propuesta con alternativas más conocidas: "la teoría de la identidad tipo-tipo", "el funcionalismo de la máquina de Turing". Según la teoría de la identidad tipo-tipo por cada término mentalista o predicado "*M*", hay algún predicado "*P*" *que se puede expresar en el vocabulario de las ciencias físicas* de manera que un ser es *M* si y únicamente es *P*. En símbolos:

[11] Lo que quiero decir es esto: los contraobjetivos entran porque el contenido es en parte una cuestión del papel *normal* o *diseñado* de un vehículo llegue éste o no a desempeñar ese papel alguna vez. Véanse Sober, 1981 y Millikan, 1984.

(1) $(x)\ (Mx \equiv Px)$

Esto es reduccionismo con una venganza, que carga con el fardo de reemplazar, en principio, todos los predicados mentalistas con predicados coextensivos compuestos funcionalmente en verdad a partir de los predicados de la física. Actualmente, hay un amplio acuerdo de que es una exigencia desesperadamente fuerte. Creer que los gatos comen pescado es, intuitivamente, un estado *funcional* que podría implementarse físicamente de distintas formas, de manera que no hay ninguna razón para suponer que el hecho común al que nos referimos en el lado izquierdo de (1) puede ser escogido con seguridad por algún predicado de la física por más complejo que éste sea. Lo que se necesita para expresar el predicado del lado derecho es, parece, un lenguaje físicamente neutral para hablar de las funciones y los estados funcionales, y los candidatos evidentes son los lenguajes usados para describir los autómatas: por ejemplo, el lenguaje de la máquina de Turing.

El funcionalista de la máquina de Turing propone entonces

(2) (x) ($Mx \equiv x$ comprende alguna k de la máquina de Turing en el estado lógico A).

En otras palabras, para que dos cosas crean que los gatos comen pescado, no necesitan ser físicamente semejantes de ninguna manera especificable pero ambas tienen que estar en un estado "funcional" especificable en principio en el lenguaje funcional más general; deben compartir la descripción de una máquina de Turing según la cual ambos están en algún estado particularmente lógico. Esta es todavía una doctrina reduccionista puesto que propone identificar cada tipo mental con un tipo funcional escogido del lenguaje de la teoría de los autómatas. Pero esto es demasiado fuerte todavía, puesto que no hay ninguna razón más para suponer que Jacques, Sherlock, Boris y Tom "tengan el mismo programa" en cualquier sentido relajado y abstracto considerando las diferencias en su naturaleza y crianza, que sus cerebros tienen alguna característica físico-química crucialmente idéntica. Debemos debilitar las exigencias para el lado derecho de nuestra fórmula todavía un poco más.

Considérese que

(3) (x) (x cree que a $p \equiv x$ se le puede atribuir predictivamente la creencia de que p).

Esto parece ser flagrantemente circular y carente de información, con el lenguaje de la derecha reflejando simplemente el lenguaje de la izquierda. Pero todo lo que necesitamos para convertir esta fórmula en una respuesta informativa, es una manera sistemática de hacer las atribuciones a las que se aludió en el lado derecho. Considérese el caso paralelo de las máquinas de Turing ¿qué tienen en común dos realizaciones o encarnaciones diferentes de una máquina de Turing cuando están en el mismo estado lógico? Simplemente esto: hay un sistema de descripción según el que ambas se describen

como realizaciones de una determinada máquina de Turing, y de acuerdo con esta descripción que predice el funcionamiento de ambas entidades, ambas están en el mismo estado de esta tabla de máquinas de Turing. Uno *no reduce* el tema de la máquina de Turing a algún modismo más fundamental; uno legitima el tema de la máquina de Turing dándole reglas de atribución y mostrando sus poderes predictivos. Si podemos legitimar de manera semejante el discurso "mentalista", no tendremos necesidad de una reducción, y esto es lo significativo del concepto de un sistema intencional. Se supone que los sistemas intencionales juegan un papel paralelo en la legitimación de los predicados mentales al papel jugado por la noción abstracta de una máquina de Turing al establecer reglas para la interpretación de artefactos tales como los autómatas computacionales. Temo que mi concepto sea penosamente informal y poco sistemático en comparación con el de la máquina de Turing, pero por otra parte el campo de acción que intenta sistematizar —nuestras atribuciones diarias en el lenguaje mentalista o intencional— es también un poco desordenado, al menos comparado con el campo claramente definido de la teoría de la función recursiva, campo de acción de las máquinas de Turing.

La analogía entre los papeles teóricos de las máquinas de Turing y los sistemas intencionales es más que superficial. Considere ese caballo de guerra de la filosofía de la mente, la Tesis de Brentano, según la cual es la intencionalidad el rasgo distintivo de lo mental: todos los fenómenos mentales muestran intencionalidad y ningún fenómeno físico la hace. Esta ha sido tomada tradicionalmente como una tesis de *irreductibilidad*: lo mental en virtud de su intencionalidad no se puede reducir a lo físico. Pero dado el concepto de un sistema intencional, podemos construir la primera mitad de la Tesis de Brentano —todos los fenómenos mentales son intencionales— como una especie de tesis *reduccionista*, paralela a la Tesis de Church en los fundamentos de las matemáticas.

De acuerdo con la Tesis de Church, todo procedimiento efectivo en matemáticas es recursivo, es decir, computable por la Turing. La Tesis de Church no se puede probar puesto que depende de la idea intuitiva e informal de un procedimiento efectivo, pero está generalmente aceptada y proporciona una reducción muy útil de una idea matemática borrosa pero útil a una noción bien definida con un alcance aparentemente igual y con mayor poder. De manera análoga, la afirmación de que todos los fenómenos mentales a los que se alude en la psicología popular es *caracterizable como sistema intencional* brindaría si fuera cierta una reducción de lo mental como se entiende ordinariamente —un campo de acción cuyos límites son fijados en el mejor de los casos por el reconocimiento mutuo y la intuición compartida— a un campo de acción claramente definido de entidades cuyos principios de organización son conocidos, relativamente formales y sistemáticos y completamente generales.[12]

[12] Ned Block (1978) presenta argumentos que supuestamente demuestran cómo las distintas teorías funcionalistas posibles de la mente caen todas en el pecado del "chauvinismo" (excluyendo inadecuadamente a los marcianos de la clase de posibles poseedores de mentes) o del "liberalismo" (incluyendo con poca propiedad distintos dispositivos, marionetas humanas y demás entre los poseedores de mentes). Mi punto de vista abarca el liberalismo más amplio, pagando con alegría el precio de algunas intuiciones recalcitrantes para la generalidad adquirida.

Este alegato de reducción, como la Tesis de Church, no se puede probar pero podría hacerse apremiante por el progreso gradual en los casos particulares (y particularmente difíciles); un proyecto que me fijo a mí mismo en otra parte (en *Brainstorms*). La tarea reductora final sería mostrar no como los términos de la teoría del sistema intencional se pueden eliminar a favor de los términos fisiológicos por la vía de la psicología cognitiva subpersonal sino casi lo contrario: mostrar cómo un sistema descrito en términos fisiológicos podría garantizar una interpretación como sistema intencional realizado.

Reflexiones: El instrumentalismo reconsiderado

"Tres clases de psicología intencional", aunque escrito antes de "Los verdaderos creyentes" lo sigue en el orden de exposición, puesto que presupone algo más de familiaridad con mi posición básica y trata algunos de los problemas con más detalle. En estas Reflexiones me concentraré en el problema de mi así llamado *instrumentalismo*, que ha causado muchas discusiones; pero antes de presentar y defender mi variedad de instrumentalismo y diferenciarlo de sus alternativas vecinas, haré un breve comentario acerca de otros cuatro temas en "Tres clases..." que proyectan sus sombras hasta otros capítulos y otras controversias.

1) ¿Cuál es la racionalidad propuesta por la adopción de la actitud intencional? Las observaciones breves y alusivas que aquí aparecen están suplementadas en detalle en el próximo capítulo: "Comprendiéndonos a nosotros mismos".

2) "La sintaxis de un sistema no determina su semántica. ¿Por medio de qué alquimia, entonces, extrae el cerebro resultados semánticamente fiables de las operaciones sintácticamente impulsadas?". Esta manera de plantear la cuestión reverberó en Searle (1980b). "El ordenador, para repetir, tiene una sintaxis pero no semántica", y de nuevo en (1982).

> En realidad Dennett no podría haber demostrado cómo pasar de la sintaxis a la semántica, de la forma al contenido mental, puesto que su clase de conductismo hace que le sea imposible aceptar la existencia de la semántica o de contenidos mentales construidos literalmente (pág. 57).

Este pasaje aclara uno de los desacuerdos fundamentales de Searle conmigo. Mientras estamos de acuerdo en que un ordenador es una máquina sintáctica, y por tanto sólo puede aproximarse al desempeño de una máquina semántica, él cree que un cerebro orgánico es un mecanismo que de alguna manera elude esta limitación de los ordenadores. El concluye su trabajo de 1980b de este modo: "Sea lo que fuere lo que el cerebro hace para producir intencionalidad, no puede consistir en ejemplificar concretamente un programa, puesto que ningún programa, por sí solo, es suficiente para la intencionalidad". Searle y yo estamos de acuerdo en que los cerebros son máquinas, pero él cree que son máquinas muy especiales:

> "¿Podría una máquina pensar?" Mi propia opinión es que *solamente* una máquina podría pensar y por cierto sólo clases muy especiales de máquinas, como el cerebro y las máquinas que tuvieran las mismas fuerzas causales que el cerebro (1980b, pág. 424).

Para mí, estas fuerzas causales conjuradas del cerebro son justo el tipo de alquimia acerca de la cual yo hacía una advertencia en mi pregunta retórica. Una máquina es una máquina, y no existe nada acerca de la construcción o de los materiales de cualquier subvariedad que pudiera permitirle trascender los límites del mecanismo y producir "verdadera semántica" además de su batido meramente sintáctico. Prosigo con más detalle este desacuerdo con Searle en los capítulos 8 y 9.

3) La sugerencia de que podemos "reducir" el concepto intuitivo de la mentalidad al de un sistema intencional, teniendo como modelo la reducción de efectividad de Turing a la computabilidad de Turing, fui yo el primero en hacerlo explícito en la Introducción a *Brainstorms*. Searle (1980b) ha llamado operacionalismo a esta manera de pensar, supuestamente una mala palabra en estos tiempos pospositivistas. Pero no para mí. Manifiesto explícitamente mi acuerdo fundamental con la opinión de Turing y hasta con su operacionalismo reputadamente chocante, en "Can Machines Think?" (1985a).

4) La distinción entre las creencias "núcleo" y las creencias virtuales o implícitas se trata con más cuidado y detalle en "Styles of Mental Representation", pero mi principal punto de vista acerca de la distinción ya fue formulado en este ensayo y todavía no ha penetrado en algunos sectores; aun cuando las consideraciones de composicionalidad o generatividad nos llevan a la conclusión de que el cerebro *tiene que* estar organizado en un conjunto modesto y explícito de elementos núcleo de los cuales "el resto" se genera de algún modo de acuerdo con la necesidad (Dennett, 1975), todavía no se ha dado ninguna razón para suponer que algunos de los elementos núcleo sean creencias mejor que algunas estructuras neurales de datos todavía innombradas e inimaginadas de propiedades ampliamente diferentes. Quienes quieren ser "realistas acerca de las creencias" —en oposición a mi instrumentalismo, por ejemplo— a menudo contestan mis argumentos acerca de la existencia inagotable de creencias diciendo que son realistas al referirse sólo a las creencias núcleo. Algunos intuyen que sólo las creencias núcleo son creencias si se habla con propiedad (por ejemplo, Goldman, 1986, págs. 201-2). ¿Qué los hace estar tan seguros de que existen las *creencias* núcleo? Soy un realista tan firme como cualquiera acerca del almacenaje de esta información de elementos núcleo en el cerebro, sea lo que fuere que resulten ser, con las cuales están ancladas nuestras interpretaciones intencionales. (Simplemente dudo, junto con los Churchland, PM 1981, 1984; PS, 1980, 1986) que esos elementos una vez individualizados sean reconocibles como las creencias que aparentamos distinguir en la psicología popular. En las reflexiones que siguen al capítulo 6 explico por qué no extraigo la moraleja de Churchland de esta duda compartida.

El instrumentalismo

> Me propongo decir que alguien es un *Realista* acerca de las actitudes proposicionales si (a) afirma que hay estados mentales cuyas ocurrencias e interacciones causan la conducta, y más aún, que lo hacen de manera que respetan (por lo menos con cierta aproximación) las generalizaciones de la psicología de la creencia/deseo del sentido común; y (b) afirma que estos mismos estados mentales causalmente eficaces son también semánticamente evaluables (Fodor, 1985, pág. 78).

Como dije en la pág. 44, soy como un realista, pero no soy un realista de Fodor puesto que espero que los verdaderos estados internos que causan la conducta no se individualizan funcionalmente, ni siquiera por aproximación, del modo en que la psicología de la creencia/deseo talla las cosas. Me he dejado llamar instrumentalista durante media docena de años (así que es culpa mía), pero no me hace nada feliz la culpa por asociación que de ahí he adquirido. Por ejemplo,

> Primera opción Anti Realista: Se podría tomar el punto de vista de un *instrumentalista* acerca de la explicación intencional... La gran virtud del instrumentalismo —acá y en cualquier otra parte— es que se reciban todas las bondades y no se sufre ningún dolor: se llega a usar la psicología de la actitud proposicional para hacer predicciones conductistas: se llega a "aceptar" todas las explicaciones, intencionales que es conveniente aceptar; pero no hay que contestar preguntas difíciles acerca de cuáles son las actitudes (Fodor, 1985, pág. 79).

El instrumentalismo clásico fue una visión abarcadora, un rechazo completo del Realismo: uno fue instrumentalista no sólo acerca de los centros de gravedad y los paralelogramos de fuerzas y el ecuador, sino también acerca de los electrones, células, planetas, todo, excepto lo que se podía observar con los sentidos desnudos. Este instrumentalismo abarcador no disculpa a sus adherentes, como dice Fodor, de tener que contestar ciertas difíciles preguntas acerca de las creencias y otras actitudes proposicionales. Desde el principio he contrastado mi Realismo acerca del cerebro y sus distintas partes, estados y procesos neurofisiológicos y mi "instrumentalismo" acerca de los estados de creencia que aparecen como *abstracta* cuando uno trata de interpretar todos esos fenómenos reales adoptando la actitud intencional. La diferencia que quise trazar es bastante conocida y (creo) polémica en otros contextos. Como observa Friedman (1981):

> La distinción entre reducción y representación no es una mera invención del filósofo; desempeña un papel genuino en la práctica científica. Los científicos mismos distinguen entre los aspectos de la estructura teórica destinados a ser tomados literalmente y los aspectos que sirven a una función puramente representativa. Nadie cree, por ejemplo, que los así llamados "espacios de estado" de la mecánica —espacio fase en mecánica clásica y espacio Hilbert en la mecánica cuántica— sean parte del mobiliario del mundo físico (pág. 4).

Mi tentativa de endosar y explotar esta distinción como una variedad del instrumentalismo *selectivo* fue evidentemente un error táctico por la confusión que causó. Debería haber abjurado del término y solamente haber dicho algo como esto: mi *ismo* es cualquier *ismo* que los realistas serios adopten con respecto a los centros de gravedad y demás, puesto que pienso que las creencias (y algunos otros aspectos mentales tomados de la psicología popular) son *así* —al ser *abstracta* más que "parte del mobiliario del mundo físico" y al ser atribuídos en declaraciones que son verdaderas sólo si las exceptuamos de cierto estándar conocido de literalidad.

Algunos instrumentalistas han respaldado el *ficcionalismo*, el punto de vista de que ciertas afirmaciones teóricas son *falsedades útiles*, y otros han sostenido que las afirmaciones teóricas en cuestión no eran ni verdaderas ni falsas sino menos instrumentos de cálculo. No defiendo ninguna de estas variedades de instrumentalismo; como lo dije cuando usé el término por primera vez antes: "la gente tiene realmente creencias y deseos en mi versión de la psicología popular, tal como tiene verdaderamente centros de gravedad". ¿Concedo entonces que las atribuciones de creencia y deseo bajo la adopción de la actitud intencional pueden ser *verdad*? Sí, pero usted me interpretará mal a menos que admita que los siguientes también son verdad.

1) La atracción gravitacional entre la Tierra y la Luna es una fuerza que actúa entre dos puntos: los centros de gravedad de los dos cuerpos.
2) Las calculadoras de mano, suman, restan, mutiplican y dividen.
3) Una Vax 11/780 es una máquina Turing universal.
4) Shakey (el "robot vidente" descrito en Dennett, 1982b) hace dibujos lineales aun cuando su CRT está apagada.

Se puede argüir que cada una de éstas sea una falsedad útil demasiado simplificada; preferiría decir que cada una es una verdad que uno debe comprender *con cierto escepticismo*. No tengo ninguna traducción oficial, canónica, de esa conocida frase, pero tampoco veo que sea necesaria. Preferiría hacer que mi punto de vista sea todo lo claro y convincente que yo pueda explicando por qué creo que todo lo que se dice acerca de la creencia tiene el mismo status (*veritas cum grano salis*, para decirlo técnicamente) como (1-4). El status deriva de los usos que encontramos para la actitud intencional.

Se puede ver a la actitud intencional como un caso limitador de la actitud del diseño: uno predice adoptando sólo una suposición acerca del diseño del sistema en cuestión: cualquiera que sea el diseño, es óptimo. Se puede ver esta suposición en acción cada vez que, en medio de la actitud de diseño propiamente dicha, un diseñador o investigador de diseño inserta un hombrecillo ingenuo (un sistema intencional como subsistema) para llenar un vacío de ignorancia. En efecto, el teórico dice, "todavía no sé cómo diseñar este subsistema, pero sé lo que se supone que debe hacer, así que finjamos que allí hay un demonio que no quiere nada más que realizar esa tarea y sabe cómo hacerlo". Uno puede entonces proseguir el diseño del sistema circundante con la suposición simplificadora de que este componente es "perfecto". Uno se pregunta cómo debe funcionar el resto del sistema, dado que su componente cumplirá con su deber.

Ocasionalmente tal esfuerzo de diseño en AI continúa literalmente instalando un módulo humano *pro tempore* para explorar las alternativas de diseño en el resto del sistema. Cuando se estaba desarrollando el sistema HWIM de reconocimiento del habla, en Bolt Beranek y Newman (Woods y Makhoul 1974), el papel del módulo de análisis fonológico que supuestamente generaba hipótesis acerca del igualmente polémico análisis de los segmentos de la entrada de datos acústicos, fue interpretado temporalmente por fonólogos humanos que miraban segmentos de espectrogramas de aserciones. Otro ser humano desempeñando el papel de módulo de control podría comunicarse con el demonio fonológico y el resto del sistema formulando preguntas y planteando hipótesis para su evaluación.

Una vez que se determinó todo lo que el resto del sistema tenía que "saber" para darle al módulo fonológico la ayuda que necesitaba, esa parte del sistema se diseño (descartando, *inter alia*, al demonio de control) y luego los fonólogos mismos pudieron ser reemplazados por una máquina: un subsistema que usó la misma entrada de datos (espectrogramas pero no visualmente codificados, por supuesto) para generar las mismas clases de dudas e hipótesis. Durante la fase de prueba del diseño, los fonólogos hicieron todo lo que pudieron para no usar todos los conocimientos extras que tenían —acerca de palabras semejantes, compulsiones gramaticales, etc.— puesto que estaban parodiando a hombrecillos *estúpidos*, a especialistas que sabían y se preocupaban únicamente por la acústica y los fonemas.

Hasta el momento en que se haga semejante esfuerzo para reemplazar el subsistema del fonólogo por una máquina, uno no está comprometido virtualmente con ninguna de las suposiciones de los tipos de diseño *acerca del funcionamiento de ese subsistema* que son genuinamente explicativos. Pero mientras tanto, se pueden hacer grandes progresos en el diseño de otros subsistemas con los que puede interactuar y el diseño del supersistema compuesto por todos los subsistemas.

El primer autómata jugador de ajedrez representado fue una broma de fines del siglo XVIII: el maniquí de madera del Baron Wolfgang von Kempelen que de verdad levantaba y movía las piezas del ajedrez jugando así una partida decente. Pasaron años antes de que se revelara el secreto de su funcionamiento: un maestro del ajedrez enano estaba escondido en el mecanismo de relojería debajo de la mesa de ajedrez y podía ver las movidas a través de los cuadros translúcidos —un homúnculo literal (Raphael 1976). Obsérvese que el éxito o el fracaso de la actitud intencional como pronosticadora es tan neutral con respecto al diseño, que no distingue el diseño del enano trabajando de, digamos, el Hitech de Berliner, un programa de ajedrez actual de considerable poder (Berliner y Ebeling 1986). Ambos trabajan; ambos trabajan bien, ambos deben tener un diseño que sea una buena aproximación a lo óptimo, siempre que lo que queramos decir por óptimo en este punto se centre únicamente en la tarea de jugar al ajedrez e ignorar todas las demás consideraciones del diseño (por ej., el cuidado y alimentación del enano *versus* el costo de electricidad; ¡pero trate de encontrar una forma eléctrica usable en el siglo XVIII!). Cualesquiera que sean sus diferencias internas, ambos sistemas son sistemas intencionales de calidad, aunque uno de ellos tenga un subsistema, un hombrecillo, que sea, él mismo, un sistema intencional to-

do lo poco problemático que se pueda encontrar.

La teoría del sistema intencional es casi literalmente una teoría de la caja negra que lo convierte en *conductista* para filósofos como Searle y Nagel, pero apenas conductista en el sentido de Skinner. Por el contrario, la teoría del sistema intencional para probar lo que Chomsky, que no es conductista, llama un modelo de la competencia, en contraste con un modelo de la actuación. Antes de que nos preguntemos cómo se diseñan los mecanismos, debemos tener claro qué es lo que se supone que harán (o podrán hacer) los mecanismos. Marr (1982) ha desarrollado mejor esta visión estratégica en sus reflexiones metodológicas acerca de su trabajo sobre visión. Distingue tres niveles de análisis. El más alto, que equivocadamente llama *computacional*, no tiene nada que ver, de hecho, con los procesos computacionales sino estrictamente (y más abstractamente) con la cuestión de a qué función está sirviendo el sistema en cuestión o, más formalmente, con qué función en el sentido matemático debe (de una u otra manera) "computar" (por ej., Newell, 1982; Dennett, 1986c, por aparecer e). A este nivel computacional uno intenta especificar formal y rigurosamente la competencia propia del sistema (Millikan, 1984, lo llamaría la función propia del sistema). Por ejemplo, uno completa los detalles en la fórmula:

"dado un elemento en el conjunto de los de '*x*' como *input*, produce un elemento en el conjunto de '*y*' como *output*, según las siguientes reglas formales..."

mientras permanece silencioso o neutral acerca de la implementación o detalle de actuación de lo que sea que haya en la caja negra competente. El segundo nivel de Marr es el nivel *algorítmico* que especifica los procesos computacionales pero es todo lo neutral que sea posible acerca de los mecanismos físicos que los implementan, que son descritos en el nivel del *hardware*.

Marr afirma que hasta que no tengamos una comprensión clara y precisa de la actividad del sistema en su nivel "computacional", el más alto, no podemos dirigir preguntas detalladas a los niveles más bajos adecuadamente ni interpretar esos datos como ya podemos haberlo hecho con los procesos al implementar esos niveles más bajos. Este es el eco de la prolongada insistencia de Chomsky sobre que el proceso diacrónico de aprendizaje de un idioma no se puede investigar totalmente hasta que no se tenga en claro el estado final de la competencia adulta hacia el que se dirige. Como el punto de vista de Chomsky, esto se entiende mejor como máxima estratégica que como principio epistemológico. Después de todo, no es imposible encontrar por casualidad algo importante en un cuadro mayor mientras uno intentaba preguntarse las cosas que terminarán siendo preguntas subsidiarias y en cierto sentido planteadas de manera ciega.

El punto de vista estratégico más revelador de Marr es que si uno tiene una visión totalmente equivocada acerca de cuál es la descripción a nivel computacional de su sistema (como, a su criterio, lo tuvieron todas las primeras teorías de la vista), sus tentativas de teorizar a niveles más bajos se verán confundidas por acertijos artefácticos espurios. Sin embargo, lo que Marr subestima, es el punto hasta el cual las descripciones a nivel computati-

vo (o actitud intencional) pueden también confundir al teórico que olvida cuán idealizadas están (Ramachandran, 1985a, b).

El hecho acerca de los modelos de la competencia que provoca mi "instrumentalismo" es que la descomposición del modelo de la competencia de alguien en partes, fases, estados, pasos o lo que sea, no *necesita* arrojar ninguna luz sobre la descomposición de las verdaderas partes mecánicas, fases, estados o pasos del sistema que se está diseñando —aun cuando el modelo de la competencia sea excelente como tal. [13]

Consideremos lo que podemos decir acerca de la competencia de dos clases muy diferentes de entidades, una calculadora manual y un captador de chistes de Newfie. Un captador de chistes de Newfie es una persona que se ríe cuando le cuentan el siguiente chiste de Newfie. (En Canadá, los chistes del habla étnica son acerca de los "Newfies —los habitantes de Terranova— y mi chiste favorito me lo contó Zenon Pylyshyn hace unos años. Cualquiera que esté familiarizado con el punto de vista de Pylyshyn acerca de la imaginería mental se preguntará si no fue arriesgado por su parte revelarme este fenómeno particular).

> Un hombre fue a visitar a su amigo el Newfie y lo encontró con las dos orejas vendadas. "¿Qué pasó?", preguntó el hombre, y el Newfie le contestó: "Me estaba planchando la camisa cuando sonó el teléfono", "eso explica una oreja, pero ¿y la otra?" "—¡Bueno, tuve que llamar a un médico!".

Si usted "entendió" el chiste, pertenece a la clase de captadores de chistes de Newfie. ¿Cómo podemos caracterizar a la capacidad necesaria para entender el chiste? Observe que el texto es radicalmente antimemático —le falta información—. Nunca habla de contestar el teléfono o de ponerse la plancha en la oreja o de la semejanza en la forma o en el peso, entre una plancha y el auricular del teléfono. Esos detalles los tiene que llenar usted solo, y tiene que tener los conocimientos necesarios, el "know-how" para poder hacerlo. Podríamos tratar de hacer una lista de "todas las cosas que tenía que saber" para captar el chiste (Charniak, 1974). Presumiblemente estas cosas son proposiciones, y confeccionaríamos una lista larguísima si nos ocupáramos de ello. Si usted, no hubiera conocido alguna de estas "cosas", no habría captado el chiste. Su conducta en esta oportunidad (suponiendo que riera, sonriera, soltara risotadas o gruñera) es una prueba casi positiva de que su caracterización intencional incluye todos estos puntos en la lista de creencias.

Al mismo tiempo, nadie tiene mucha idea de cómo sería una buena teoría que procesara su comprensión del cuento. La mayor parte de los lectores diría, probablemente, que lo que las pasó mientras leían el chiste fue que se

[13] Bechtel (1985) observa: "Dennett parece reducir la cuestión de instrumentalismo *versus* realismo a un punto empírico acerca de cómo está estructurado el sistema cognitivo humano: si resulta que hay un razonable paso de los modismos intencionales a los estados de procesamiento, entonces el realismo quedará reivindicado, mientras que el instrumentalismo se justificará si no se produce ese cambio" (pág. 479).

formaron una imagen visual de cierto tipo en el ojo de la mente —primero la del Newfie llevando la mano hacia el teléfono y luego levantando ambas manos a la vez— en cuanto leyeron la parte acerca del llamado telefónico. Es bastante fácil contar un cuento introspectivo, sincero y plausible de esta clase, pero si les pidiéramos a muchos lectores que compartieran con nosotros la fenomenología que recordaron, no nos sorprendería que hubiera una considerable variación acerca de los temas centrales esenciales. Y aun si alguien negara haber experimentado imaginería alguna, pero que sin embargo, se había reído mucho (y por cierto que hay gente definidamente así en el muestreo), supondríamos que una historia de procesamiento *inconsciente* —por otra parte casi con el mismo contenido— debe ser cierto con respecto a ese individuo. Después de todo no es magia y en cada individuo debe haber un proceso sensible a la información que lo lleva de las palabras leídas a la risa ahogada.

La historia a contar acerca del proceso real en individuos diferentes, no la contará por la teoría del sistema intencional. Entre las proposiciones que uno tiene que creer para entender el chiste, está la proposición de que la gente habitualmente contesta el teléfono cuando éste suena. Otra es que contestar el teléfono implica levantar el auricular con una mano y ponerlo en contacto con la oreja. No los aburriré con una enumeración de otras creencias necesarias. Estas creencias no sólo no "vienen a la mente" en forma de juicios verbales (que podrían distraerlo a uno seriamente de prestar atención a las palabras del cuento), sino que también es sumamente poco plausible suponer que todas y cada una de ellas sea consultada, independientemente, por un mecanismo computacional diseñado para entretejer las lagunas del cuento mediante un proceso de generación deductivo. Y sin embargo, la información expresada en esas oraciones tiene que estar de algún modo en la cabeza de aquellos que entienden el chiste. Podemos predecir y explicar algunos fenómenos a este nivel. Por ejemplo, que en esta época de ropas que se lavan y no se planchan, hay una generación de niños que están creciendo, que incluye a muchos que en realidad no saben cómo es la acción de planchar una camisa; estos individuos protegidos se sentirán desconcertados por el chiste puesto que carecen de algunas creencias esenciales.

La lista de creencias nos brinda una buena idea general de la información que debe estar en la cabeza, pero si la consideramos como una lista de axiomas de los cuales un proceso derivativo deduce el "quid del chiste" podemos tener esbozos de un modelo de la actuación, pero es un modelo de la actuación particularmente feo. De manera que aunque está lejos de ser ocioso catalogar el conjunto mínimo de creencias del captador canónico de chistes de Newfie, esa caracterización de la actitud intencional, con todo su poder predictivo y hasta explicativo, no arroja habitualmente ninguna luz sobre los mecanismos subyacentes.

Considere ahora el conocido modelo de competencia de una calculadora de mano. Supongamos que utilizo mi habilidad para hacer una aritmética de lápiz y papel para predecir el comportamiento de mi calculadora de bolsillo (a los filósofos les gusta a menudo complicarse la vida). El formalismo de la aritmética resulta un "nivel computacional" ejemplar: riguroso y preci-

so hasta la exageración, provee de normas para predecir exactamente qué *output* entregará la caja negra para cualquier *input*. Estas normas transforman el *input* ("26 x 329") a través de una serie de fases (6 veces 9 es 54, anota el 4, llévate el 5...") hasta que eventualmente se produce un *output*, el cual (si hemos hecho bien la computación) se confirma o se prueba por el comportamiento de la calculadora. ¿Pero las etapas componentes del proceso por el cual yo calculé mi predicción reflejan algunas etapas componentes del proceso por el cual la calculadora produjo su resultado pronosticado? Unas sí y otras no. Aquellas que sí lo hacen en una ocasión, no tienen necesariamente que hacerlo en otra; el "ajuste", cuando existe, puede ser accidental. Hay cantidades infinitas de maneras de manejar una calculadora para que se ajuste a nuestro modelo de competencia y sólo algunas de ellas "se parecen mucho" a los detalles de los cálculos del modelo de competencia.

Por supuesto que no existe un modelo aritmético de competencia único para mi calculadora de mano. La aritmética deconoce la finitud, pero mi calculadora no. A su manera finita, utiliza un proceso que implica aproximación y/o el truncamiento (dos tercios resulta, 6666) o el redondeo (dos tercios son, 6667). La predicción y el experimento determinarán cuál. La aritmética dice que 10 dividido por 3, multiplicado por 3 es 10, y que 20, dividido por 3, multiplicado por 3, es 20, pero mi calculadora dice que la primera respuesta es 9.9999999, y la segunda es 19.999999. Descubrir este desajuste entre la perfección y la realidad nos proporciona una pista poderosa acerca de la verdadera maquinaria de la calculadora.

Aquí puedo usar el modelo de competencia inicial para generar hipótesis cuyas falsificación arroja luz sobre la verdadera organización de la maquinaria. esto da lugar a un refinamiento del modelo de competencia mismo, en el sentido de convertirlo en un modelo de actuación. Preguntar cuál de los algoritmos posibles para la aritmética aproximada es el que usa la caculadora, es descender al nivel algorítmico. Si, cuando descendemos al nivel algorítmico, perdemos algunas de las categorías conocidas del nivel computacional (quizá nada cuente tanto como el paso donde usted "escribe el cuatro y se lleva el cinco"), esto no disminuirá la utilidad práctica de usar el modelo de competencia puro como algo que predice o como un ideal, un especificador de lo que el sistema debería hacer, pero sí que nos impedirá descubrir la identidad (material) de aquellos que parecen ser, desde esta perspectiva, los componentes y fases del proceso. [14]

Alejarse de un ideal exageradamente simple hacia un mayor Realismo no es siempre una táctica prudente. Depende de lo que usted quiera; a veces una predicción rápida y burda es mejor que una extensión de la comprensión científica detallada hasta en la ciencia. El hecho de que se pueda esperar que un objeto se acerque con garantías a lo óptimo (o a la racionalidad) puede ser

[14] ¡Por favor, obsérvese que ***nuestra*** capacidad de "hacer aritmética" sin sucumbir directamente al error del truncado o el redondeo, no prueba en absoluto que no seamos mecanismos o que no seamos finitos! Tal vez seamos mecanismos finitos *más grandes y lujosos*. Hay sistemas de computación, tales como MACSYMA que son expertos en manipulaciones algebraicas y tienen modos alternativos, por ejemplo, de representar los números irracionales.

un hecho más profundo y valioso que cualquier otro que se pueda obtener desde el punto de vista más realista y más detallado.

Siempre he destacado el poder predictivo real de la actitud intencional pura. He manifestado, por ejemplo, que uno puede usar la actitud intencional para predecir la conducta de un rival (humano o un artefacto) desconocido en una partida de ajedrez, y esto a veces provoca la objeción (por ej., Fodor, 1981, capítulo 4) de que lo que hace del ajedrez un juego interesante es precisamente que los movimientos del oponente *no* son predecibles desde la actitud intencional. Si lo fueran, el juego sería aburrido.

Esta objeción coloca los estándares de predicción demasiado alto. En primer lugar, están esas situaciones durante la última parte de la partida cuando hablamos de "movimientos obligatorios", y éstos son predecibles desde la actitud intencional (y sólo desde ella) con un virtual 100 % de fiabilidad. (La única fuente de dudas que uno tiene es si, o cuándo, el contrario va a renunciar al siguiente "movimiento obligatorio" o va a efectuarlo.) Los movimientos obligatorios son obligatorios sólo en el sentido de que están "dictados por la razón". Un contrario irracional (ya sea humano o un artefacto) podría autoderrotarse de manera contraproducente al no efectuar un movimiento obligatorio, y en cuanto a eso, un oponente racional (una vez más un ser humano o un artefacto) podría tener un motivo ulterior superior a evitar la derrota en el ajedrez.

¿Y qué hay de los movimientos del contrario en la mitad de la partida? Estos se pueden predecir rara vez con fiabilidad hasta la singularidad (y, como lo afirman los críticos, eso es lo que hace interesante el ajedrez) pero es una situación rara cuando los treinta o cuarenta movimientos *legales* que tiene el contrario no pueden ser reducidos por la actitud intencional a una corta y desordenada lista de media docena de movimientos más probables por los que se puede apostar con éxito si se recibiera dinero constante por todos los movimientos legales. Esta es una ventaja predictiva tremenda, sacada del aire transparente frente a una ignorancia casi total de los mecanismos que intervienen gracias al poder de la actitud intencional.

Es así como los programas de ajedrez están diseñados para economizar; en lugar de dedicarle una atención igual a toda las continuaciones posibles de la partida, en algún punto un programa de ajedrez se concentrará en aquellas ramas del árbol de la decisión en las cuales el contrario ejecuta (lo que el programa calcula que será) *su* mejor respuesta. No establece ninguna diferencia para el programa de ajedrez que su contrario sea humano o un artefacto; calcula simplemente a partir de la suposición de que cualquier contrario con el que valga la pena jugar tratará de hacer las mejores jugadas que pueda. Los errores "temerarios" pueden ser así maneras de sorprender —de alterar las expectativas— de un programa que economiza de manera demasiado optimista con respecto de sus contrarios. ¿Habría que re-diseñar el programa para estar a la expectativa de modelos de esa elección "subóptima" en sus rivales como un paso en la dirección de un mayor realismo, una comprensión más perfilada de sus oponentes individuales? Tal vez. Todo depende de los gastos de mantenimiento, y de acceder a ese detalle bajo la presión del reloj. Probablemente los riesgos de ser engañados por "errores" deliberados son tan bajos que son un precio tolerable por la velocidad y efi-

ciencia de usar las presunción menos realista de lo óptimo. A Fodor (1981b) se le escapa esto cuando sugiere que la presunción de racionalidad impediría que un jugador de ajedrez planeara una amenaza.

> Por tanto, si asumo la racionalidad de Black, supongo *inter alia*, que Black nota la amenaza. De donde, lo que espero y predigo es precisamente que la amenaza pasará inadvertida. Al predecir esto, no he abandonado la actitud intencional. Por el contrario, puedo predecir racionalmente el deliz de Black precisamente *a causa de lo que sé o creo acerca de los estados intencionales de Black: en especial acerca de lo que es probable que él note o que no note* (pág. 108).

¿Qué es lo que es posible que Black no note? Puedo decírselo sin saber siquiera quién es Black. Al ser aproximadamente racional, no es probable que Black note amenazas que tomaría mucho tiempo y esfuerzo descubrir y es sumamente probable que note las amenazas evidentes. Si como Fodor supone, es poco probable que Black se dé cuenta de la amenaza, debe ser porque la amenaza está algo distante en el árbol de la búsqueda, y por lo tanto, puede muy bien quedar fuera del foco de atención más o menos óptimo de Black.

De manera que aun cuando estamos planeando explotar las debilidades de otra gente racional, hacemos uso de la presunción de la racionalidad para guiar nuestros esfuerzos (véase Dennett, 1976, reimpreso en *Brainstorms*). Puesto que el poder general de la actitud intencional no se explica así por medio de ningún conocimiento que pudiéramos tener acerca de los mecanismos de los objetos que de tal modo comprendemos, continúo resistiendo al estigma de Realismo que saca esa conclusión del éxito diario de la actitud de que deben haber estados como creencias y como deseos en todos los objetos semejantes. Bechtel (1985) ha sugerido, constructivamente, que esto aun no me excluye de una variedad de Realismo. Podría reconstruir mi instrumentalismo como Realismo acerca de ciertas propiedades abstractas de la relación: propiedades que relacionan un organismo (o un artefacto) con su entorno de determinadas maneras indirectas. Millikan (1984) tiene una explicación positiva del mismo tipo. Hasta donde puedo ver, ésta es por cierto una opción ontológica viable para mí, pero como lo demostrará el capítulo 5, por las propiedades acerca de las cuales uno terminaría siendo un realista apenas valen la pena de hacer el esfuerzo. [15]

[15] Consideré un movimiento táctico similar en el Capítulo 1 de ***Content and Consciousness***, donde la opción era una teoría de identidad igualmente tortuosa y me decidí en su contra: "¿No ha perdido uno el meollo de la teoría de la ***identidad*** una vez que comienza a tratar oraciones enteras como nombres de situaciones o estado de cosas que luego se proclaman idénticas a otras situaciones o estados de situaciones?" (pág. 18n).

4

Comprendiéndonos a nosotros mismos

En "Dennett on Intentional Systems" (1981), Stephen Stich da una explicación vivaz, simpática y generalmente exacta, con objeciones y contraposiciones detalladas (véanse también Stich, 1980 y Dennett, 1980c), de mi punto de vista. Mi refinamiento propuesto de la idea popular de la creencia (por la vía del concepto de un *sistema intencional* "nos dejaría incapacitados para decir gran parte de lo que ahora deseamos decir acerca de nosotros mismos". Para que ésta sea una objeción, él debe de querer decir que nos dejaría incapaces de decir gran parte de lo que *precisamente* queríamos decir, porque presumiblemente es cierto. Debemos ver entonces qué verdades supone él que quedan fuera de nuestro alcance en mi explicación. Muchas de ellas mienten, dice, en el reino de los hechos acerca de nuestras faltas cognitivas que, de acuerdo con mi explicación, no pueden tener ninguna descripción cognitiva: "Si tratamos de traficar con las nociones de creencia y deseos del sistema intencional, entonces, simplemente, no podríamos decir todas esas cosas que necesitamos decir acerca de nosotros mismos y nuestros amigos cuando nos ocupamos de la idiosincrasia, fallos y crecimiento cognitivo de cada uno" (pág. 48). El da verdaderos ejemplos. Entre ellos están el astronauta olvidadizo, el chico del puesto de limonada que da mal el cambio y el hombre que ha calculado mal el saldo de su cuenta. Estos tres son casos de fracaso cognitivo simple y poco misterioso —casos de gente que *comete errores*— y Stich afirma que mi punto de vista no los puede incluir. Algo que llama la atención en los tres casos es que, a pesar de la expresión sumaria de su objeción, éstos no son casos de irracionalidad conocida o casos de fracasos deductivos. No son casos de lo que comúnmente llamaríamos irracionalidad, y puesto que hay casos sumamente apremiantes de lo que ordinariamente *llamaríamos* irracionalidad (y puesto que Stich los conoce y cita, por cierto, algunos de los casos mejor documentados [Wason y Johnson-Laird, 1972; Nisbett y Ross, 1980]), vale la pena preguntar por qué cita en cambio estos casos de cálculos equivocados como prueba en contra de mi punto de vista. Haré esta pregunta en pocas palabras, si bien primero debería conceder que en cualquier caso éstos son ejemplos de conducta subóptima de la clase que mi punto de vista se supone que no puede manejar.

Sostengo que esos errores, ya sea como *funciones defectuosas* o los productos del error de diseño son imprevisibles desde la actitud intencional, una afirmación con la que Stich podría estar de acuerdo, pero continuo para

afirmar que habrá inevitablemente estabilidad o aspecto problemático en la mera *descripción* de tales deslices en el sistema intencional, al nivel en el cual se atribuyen las creencias y deseos de la gente; y aquí parece, en un principio, que Stich debe de tener razón. Porque aunque pocas veces suponemos que podemos *predecir* determinados errores de la gente desde nuestra perspectiva de psicología popular ordinaria, no parece haber nada más directo que la descripción psicológica popular de esos casos conocidos. Presumiblemente, ésta sea parte de la razón por la cual Stich eligió estos casos: no son nada polémicos.

Miremos sin embargo, más de cerca, uno de los casos agregando más detalles. El cartel del chico dice: "LIMONADA-12 centavos el vaso". Le doy 25 centavos, él me da un vaso de limonada y luego me da 11 centavos de cambio. Ha cometido un error. Ahora bien, ¿qué podemos esperar de él cuando le señalamos su error? Que muestre sorpresa, que se ruborice, que se golpee la frente, que se disculpe y que me dé dos centavos. ¿Por qué esperamos que demuestre sorpresa? Porque le atribuimos la creencia de que me ha dado bien el cambio, le sorprenderá saber que no lo ha hecho (véase Weizenfeld, 1977). ¿Por qué esperamos que se ruborice? Porque le atribuimos el deseo de no estafar (o de ser visto estafando) a sus clientes. ¿Por qué esperamos que se golpee la frente o reconozca de alguna otra manera su falta? Porque le atribuimos no sólo la creencia de que 25 − 12 = 13, sino también la creencia de que eso es evidente, y la creencia de que nadie de su edad debería cometer errores así. Aun cuando no podemos prever este error especial —aunque podamos haber hecho una predicción estadística de que probablemente cometería algún error así antes de que acabara el día—, podemos recoger la madeja de nuestra interpretación intencional una vez que él haya cometido su error y predecir sus reacciones y actividades ulteriores con nada más que el riesgo concomitante habitual. Parece entonces, a primera vista, que la atribución de deseo en este caso es tan fácil, predictiva y estable como lo es siempre. Pero miremos aun más cerca. Es verdad que el muchacho ha cometido un error, ¿pero *exactamente qué error?* Todo esto depende, por supuesto, de cómo contamos el cuento, hay muchas posibilidades diferentes. Pero sea cual fuere el cuento que contemos destaparemos un problema. Por ejemplo, podríamos suponer plausiblemente que hasta donde llega nuestra evidencia actual el chico cree:

1) que me ha dado bien el cambio
2) que le di un cuarto
3) que su limonada cuesta 12 centavos
4) que un cuarto es 25 centavos
5) que el "*dime*" es una moneda de 10 centavos
6) que el penique es 1 centavo
7) que me dio un "dime" y un penique de cambio
8) que 25 − 12 = 13
9) que 10 + 1 = 11
10) que 11 ≠ 13

Sólo (1) es una creencia falsa, ¿pero cómo se puede decir que él cree *esa* si cree todas las demás? Seguramente no es plausible alegar que él ha ***deduci-***

do mal (1) de cualquiera de los otros, directa o indirectamente. Es decir, no nos sentiríamos inclinados a atribuirle la inferencia de (1) directamente de (7) y que tal vez él deduciría:

11) que me dio 11 centavos de cambio

de (9) y (7) —debería hacerlo después de todo— pero *no tendría sentido* suponer que *infirió* (1) de (11) a menos de que estuviéramos en un mal entendido

12) de que 11 centavos es el cambio correcto de un cuarto.

Esperaríamos que él creyera *eso* si él creyese (13)

13) que 25 − 12 = 11

y podríamos haber contado esta historia de manera que el chico tenía simplemente esta falsa creencia —y no creía en (8)— (podemos imaginar por ejemplo que eso es lo que el padre le dijo cuando él se lo preguntó, esto nos daría un caso que no sería para nada plausible de irracionalidad o ni siquiera de error de cálculo, sino simplemente un caso de un pensador perfectamente racional con una única creencia falsa [que luego genera entonces otras creencias falsas como (1)]. Correctamente Stich no quiere considerar semejante caso, ya que por supuesto yo reconozco la posibilidad de una mera creencia falsa cuando se pueden contar historias acerca de su adquisición. Si atribuimos (13) y *retenemos* (8) obtenemos un caso flagrante y raro de irracionalidad: alguien que cree simultáneamente que 25 − 12 = 13, 25 − 12 = 11 y 13 = 11. Esto no es para nada lo que habíamos supuesto, pero es tan raro que estamos obligados a encontrar las atribuciones conjuntas francamente increíbles. Algo tiene que ser. Si decimos, como propone Stich, que el muchacho todavía "no suma muy bien mentalmente", ¿cuál es la implicación? ¿Que no cree verdaderamente en la inconsistente tríada, que él más bien comprende las nociones aritméticas lo suficientemente bien como para tener las citadas creencias? Esto es, si decimos lo que dice Stich y atribuimos *también* las creencias inconsistentes, todavía tenemos el problema de la irracionalidad bruta, demasiado absoluta para tolerarla; si tomamos la observación de Stich en serio o retiramos la atribución, entonces Stich está coincidiendo conmigo: aun los errores más simples y conocidos nos exigen que recurramos a cuotas de susto u otras advertencias acerca de la verdad literal del conjunto total de atribuciones.

Hay algo obtuso, por supuesto, acerca de la indagación previa en favor de un juego de creencia total que rodea el error. La exigencia de que descubramos una inferencia —aunque sea una inferencia equivocada— a la creencia falsa (1) es la exigencia de que encontremos una práctica o tendencia con algo parecido a una razón de ser, el ejercicio de la cual ha llevado en este caso a (1). Ninguna mera sucesión en el tiempo, ni siquiera la causalidad regular es suficiente en sí misma para ser tenida en cuenta como deducción. Por ejemplo, si supiéramos que el muchacho fue llevado directamente de su creencia (6) de que 1 penique es 1 centavo a su creencia (2) de que le di un cuarto, no importa cuán habitual e inevitable sea su paso por (6) y (2), no lo llamaríamos *inferencia*. Las inferencias son pasajes del pensamiento para los cuales hay una razón, pero la gente no comete errores por culpa de razones. Las razones exigentes (como contrarias a las "meras" causas) de los errores generan estructuras espurias de creencia, como lo acabamos de ver en (11 —

13), pero consentir simplemente en la atribución de la creencia sin sentido no es mejor. No es como si *nada* llevara al muchacho a creer (1); no es como si esa creencia fuera totalmente infundada. No suponemos, por ejemplo, que hubiera creído (1) si hubiera tenido la mano vacía o llena de cuartos o si yo le hubiera dado un dólar o una tarjeta de crédito. El basa de algún modo su creencia errónea en una percepción distorsionada, confusa o equivocada de lo que me está dando, de lo que le di y las relaciones apropiadas entre todo eso.

El chico está básicamente en la cima de la situación y no es un mero robot expendedor de cambio; no obstante, debemos descender del nivel de las creencias y los deseos a algún otro nivel de teoría para describir su error, puesto que ninguna explicación en términos de sus creencias y deseos tendrá del todo sentido completamente. En algún punto nuestra explicación tendrá que habérselas con la completa insensatez de la transición en cualquier error.

Mi análisis tal vez tendencioso de un único ejemplo constituye a duras penas un argumento en favor de mi afirmación general de que éste será siempre el resultado. Está presentado como un desafío: trate usted mismo de contar la historia total de creencia que rodea un error tan simple, y vea si no descubre la misma disyuntiva que he ilustrado.

Los errores del tipo expuesto en este ejemplo son deslices de los buenos procedimientos, no manifestaciones de fidelidad a un mal procedimiento o principio. La confirmación parcial de nuestra hipótesis de trabajo ineludible de que el muchacho es fundamentalmente racional es su ruboroso reconocimiento del error. No defiende su acción cuando se la señala, sino que voluntariamente corrige su error. Esto representa un constraste llamativo con la conducta de los agentes en los casos putativos de irracionalidad auténtica mencionados por Stich. En estos casos, la gente no sólo persiste en sus "errores" sino que también defiende obstinadamente su práctica, y hasta encuentra defensores entre los filósofos (véase Cohen, 1981). Por lo menos *no es evidente* que haya casos de conducta o pensamiento sistemáticamente irracionales. Los casos que han sido propuestos son todos polémicos, que es exactamente lo que mi punto de vista predice: no existe tal cosa como un caso de "irracionalidad conocida" convenida u obvia. Esto no quiere decir que seamos siempre racionales, sino que cuando no lo somos, los casos desafían la descripción en los términos ordinarios de creencia y deseo. No hay ningún misterio acerca de *por qué* esto tiene que ser así. Una interpretación intencional de un agente es un ejercicio que intenta *encontrarles sentido* a los actos de la gente, y cuando ocurren actos que no tienen sentido, no pueden ser interpretados directamente en los términos del "tener sentido". En algo hay que ceder: concedemos que el agente en "cierto modo" cree esto o aquello, o cree esto o aquello "la mayor parte de las veces", o cree en alguna falacia que crea un contexto en el cual lo que había parecido ser irracional resulta después de todo ser racional (véanse, por ejemplo, las sugerencias de Cohen, 1981). Estas actitudes de retirada están ellas mismas sujetas a los tests habituales acerca de la atribución de deseos, de manera que encontrar meramente una actitud de retirada no significa confirmarla. Una vez en disconformidad, la búsqueda continúa al encuentro de otra interpretación salvadora. Si

no hay ninguna interpretación salvadora —si la persona en cuestión es irracional— no se transará con ninguna interpretación.

La misma retirada del abismo se encuentra en los casos simples de errores de cálculo o equivocaciones que Stich nos recuerda, pero con algunos trucos agregados dignos de tener en cuenta. En el caso del vendedor de limonada, podríamos excusarnos de hacer nuevos intentos de separar sus creencias, admitiendo simplemente que si bien él conocía (y por lo tanto creía en) todos los datos reales, "olvidó" o "pasó" por alto temporalmente algunos de ellos, hasta que se los recordamos. Esto tiene la apariencia de ser una modesta hipótesis psicológica pequeña: algo aproximadamente en el sentido de que aunque una cosa u otra estaba guardada a buen recaudo dentro de la cabeza del agente, donde le correspondía, sus señas estaban temporalmente traspapeladas. Semejante historia puede muy bien estar sustentada al final dentro de una teoría psicológica confirmada y detallada (véase Cherniak, 1983, 1986; Thomason, 1986), pero es importante observar que en la actualidad formulamos estas hipótesis simplemente sobre la base de nuestro aborrecimiento por el vacío de la contradicción.

Por ejemplo, piense en la distracción, un mal muy bien denominado, parece. A la hora del desayuno se me recuerda que hoy voy a jugar al tenis con Paul en lugar de almorzar. A las 12.45 me encuentro liquidando el postre cuando Paul, en ropa de tenis, aparece a mi lado y me hace acordar de golpe. "¡Se me olvidó por completo!", declaro, ruborizándome ante mi propia distracción. ¿Pero por qué digo *eso*? ¿Es porque, según lo recuerdo, ni un solo pensamiento consciente acerca de mi encuentro tenístico me pasó por la cabeza después del desayuno? Eso podría ser verdad, pero quizás ningún pensamiento consciente de que hoy iba a almorzar se me ocurrió tampoco en el ínterin, y aquí estoy, terminando de almorzar. Tal vez si yo *hubiera* pensado concretamente en ir a almorzar como de costumbre, ese mismo pensamiento me hubiera hecho recordar que, en realidad, no iba a almorzar. Y en cualquier caso, aun si ahora recuerdo que *sí* se me ocurrió a media mañana que hoy iba a jugar al tenis —sin ningún provecho, evidentemente— seguiré diciendo que después se me olvidó.

¿Realmente por qué estoy ansioso por insistir en que lo olvidé por completo? ¿Para asegurarle a Paul que no lo dejé plantado a propósito? Tal vez, pero eso debería ser lo suficientemente claro como para no necesitar decirlo, y si mi ansiedad es por no querer ofenderlo, no estoy logrando del todo mi propósito, puesto que no es nada halagador haber sido olvidado tan por completo. Creo que un motivo primordial para mi afirmación es simplemente descartar la posibilidad de que de otro modo surgiría: soy gravemente irracional; creo estar jugando al tenis a la hora del almuerzo y que estoy libre para ir a almorzar como de costumbre. No puedo actuar según ambas creencias a la vez; según cualquiera de las dos que actúe, declaro que la otra se me olvidó. No según alguna evidencia introspectiva (puesto que, después de todo, puedo haber pensado varias veces en el tema en el período intermedio relevante), sino basándome en *principios generales*. No importa cuán cerca del mediodía pude haber pensado en mi cita tenística; si termino almorzando como de costumbre, esa cita se me *debe de* haber olvidado a último momento.

No hay ninguna relación directa entre nuestro pensamiento consciente y las ocasiones en que decimos que nos olvidamos de algo. Supongamos que alguien me invita a almorzar hoy y contesto que no puedo: tengo otro compromiso, pero por mi vida que no recuerdo cuál es, me acordaré más tarde. Aquí, aunque en algún sentido, he olvidado mi compromiso tenístico, en otro no me sucedió, puesto que mi creencia de estar jugando al tenis, aun cuando no sea consciente recuperable (por el momento), ya está haciendo algo por mí: me está impidiendo acudir a la cita conflictiva. Subo a mi auto y llego a la bocacalle: la izquierda me lleva a casa para almorzar; la derecha, a la pista de tenis; esta vez giro a la derecha sin la ventaja de un pensamiento consciente que me acompañe en el sentido de que hoy juego al tenis a la hora del almuerzo. No lo he olvidado, sin embargo; si lo hubiera olvidado sin duda habría girado a la izquierda (véase Ryle, 1958). ¡Hasta es posible que nos olvidemos de algo mientras pensemos conscientemente en ello! "Ten cuidado con esta cacerola, 'digo', está muy caliente" —mientras estiro la mano y me quemo con la misma olla de la que estoy hablando. El colmo de la distracción, sin duda, pero posible. Indudablemente diríamos algo así como "¡No pensaste en lo que decías!"— lo que no significa que las palabras salieran de mi boca como de la de un autómata, pero que si hubiera creído *verdaderamente* en lo que estaba diciendo, *no podría* haber hecho lo que hice. Así, si puedo no pensar lo que estoy diciendo, también podría en un caso tan raro como éste, no pensar en lo que estaba pensando. Podría pensar "cuidado con esa olla caliente", *para mis adentros* al mismo tiempo que pasaba por alto la advertencia.

Se tiene cierta tentación de decir que en ese caso, aunque yo sabía muy bien que la olla estaba caliente, simplemente lo olvidé por un momento. Tal vez queremos reconocer esta clase de olvido, pero obsérvese que no es para nada la clase de olvido que se supone que ocurre cuando decimos: "He olvidado el número de teléfono de la compañía de taxis a la que llamé hace dos semanas" o "he olvidado la fecha del cumpleaños de Hume". En esos casos suponemos que la información desapareció para siempre. Los recordatorios y las insinuaciones no me van a ayudar a recordar. Cuando yo digo "me olvidé completamente de nuestra cita para jugar al tenis", no quiero decir que la olvidé completamente, como lo demostraría si al llegar Paul en ropa de tenis yo hubiera estado completamente desconcertado por su presencia y negado todo recuerdo de haber concertado la cita.

Algunas otras locuciones conocidas de la psicología popular pertenecen a la misma familia: "notar", "pasar por alto", "ignorar" y hasta "llegar a la conclusión". La impresión inicial que uno tiene es que aplicamos estos términos a nuestros propios casos sobre la base de la instrospección directa. Es decir, que clasificamos distintos actos conscientes propios como conclusiones, observaciones y demás, ¿pero qué pasa con ignorar o pasar por alto? ¿Nos encontramos a nosotros mismos haciendo esas cosas? Sólo retrospectivamente, de un modo autojustificativo y autocrítico: "Yo ignoré la jugada de los peones del lado de la reina", dice el ajedrecista, "porque estaba tan claro que el movimiento importante involucraba a los caballos del lado del rey". Si hubiera perdido la partida habría dicho: "yo simplemente no me fijé en la juga-

da de los peones del lado de la reina, puesto que estaba bajo el error de que el ataque del lado del rey era mi único problema".

Suponga que alguien pregunte: "¿Te diste cuenta de qué manera Joe eludía tus preguntas ayer?". Yo podría contestar "sí", aunque por cierto *no tuve ningún pensamiento consciente* (que pueda recordar) en ese momento, acerca de cómo Joe estaba eludiendo mis preguntas; sí puedo, sin embargo, ver que mis reacciones hacia él (tal como las recuerdo) tomaron en cuenta su actitud evasiva, declararé (con justicia) que no lo noté, puesto que hice lo adecuado a las circunstancias, debo haberlo notado, ¿no?

Para que en este momento ustedes capten la esencia de mi relato de distracción, tuvieron que concluir de mi observación acerca de "liquidarme el postre" que acababa de terminar el almuerzo y que había olvidado mi compromiso tenístico. Seguramente llegaron a esa conclusión, ¿pero lo hicieron conscientemente? Algo remotamente parecido a: "¿Mmm, debe de haber almorzado..." le pasó por la cabeza? Probablemente no. No es más probable que el muchacho vendedor de limonada pensara conscientemente que los 11 centavos que tenía en la mano eran el vuelto correcto. "Pues bien, si no lo pensó conscientemente, lo hizo inconscientemente; debemos formular un pensamiento inconsciente de control en ese sentido para explicar, o fundamentar, o *ser* (!) su creencia de que está dando el vuelto correcto.

Es tentador suponer que cuando nos apartamos del abismo de la irracionalidad y encontramos un distinto nivel de explicación con el cual enriquecer nuestra descripción de los errores (o en cuanto a eso de pasajes del pensamiento sumamente oportunos) la arena a la cual legamos es la arena de la psicología popular de los pensamientos, las conclusiones, los olvidos, etc..., no meros *estados* mentales abstractos como la creencia, sino episodios, actividades o procesos concretos y mensurables en el tiempo que pueden ser modelados por constructores psicológicos de modelos y medidos y probados directamente en los experimentos. Pero como lo sugieren los ejemplos recién discutidos (aunque de ninguna manera lo prueban), sería poco prudente que modeláramos nuestra psicología académica; sería demasiado cerca de estas *illata* putativas de la teoría popular. Postulamos todas estas actividades y procesos mentales aparentes para encontrarle sentido a la conducta que observamos, para, en realidad, encontrarle el mayor sentido posible a la conducta, especialmente cuando la conducta que observamos es la nuestra propia. Los filósofos de la mente solían desviarse de su camino para insistir en que el acceso que uno tiene a su propio caso en esos temas es muy diferente al acceso que uno tiene a los de otros, pero a medida que aprendemos más acerca de las distintas formas de la psicopatología y hasta de las debilidades de personas aparentemente normales (véanse Nisbett y Wilson, 1977) se vuelve más plausible suponer que aunque todavía hay algunos rincones de privilegio indisputable, algunos temas en los cuales nuestra autoridad es invencible, cada uno de nosotros es, en muchos aspectos, una especie de autopsicólogo inveterado, que *inventa sin esfuerzo* interpretaciones intencionales de nuestras propias acciones en una mezcla inseparable de confabulación, autojustificación retrospectiva, y (a veces, sin duda) buena teorización. Los casos llamativos de confabulación por sujetos que están bajo hipnosis o que

sufren de distintos bien documentados desórdenes cerebrales (el síndrome de Korsakoff, el cerebro dividido, distintas "agnosias") plantean la perspectiva de que tales despliegues de virtuosismo de autointerpretaciones completamente carentes de apoyo, no son manifestaciones de una habilidad repentinamente adquirida como respuesta ante un trauma, sino de una forma normal desenmascarada (véanse Gazzoniga y Ledoux, 1978; también Gardener, 1976, para las explicaciones gráficas de tales casos).

Como productos de nuestros propios esfuerzos por encontrar el sentido de nosotros mismos, las actividades mentales putativas de la teoría popular no son precisamente el terreno neutral de hechos y procesos al que podemos recurrir en busca de explicaciones cuando las exigencias normativas de la teoría del sistema intencional se enredan con algo de irracionalidad. Tampoco podemos suponer que sus equivalentes en una psicología cognitiva desarrollada, ni que a sus "realizaciones" en la estructura húmeda del cerebro les irá mejor.

Stich mantiene la imagen de una psicología completamente libre de pautas, naturalizada, que pueda *resolver* las indefiniciones de la teoría del sistema intencional apelando, finalmente, a la presencia o ausencia de estados y sucesos verdaderos, funcionalmente salientes, causalmente potentes que se pueden identificar y a los que se les puede atribuir contenido independientemente de los cánones problemáticos de racionalidad ideal que mi punto de vista exige. ¿Qué creía realmente el vendedor de limonada? ¿O, en cualquier caso, cuál era el *contenido exacto* de la secuencia de estados y sucesos que figuran en la descripción cognitiva de su error? Stich supone que en principio podremos decirlo, aun en los casos donde mi método aparece con las manos vacías. Yo afirmo, por el contrario, que al igual que la interpretación de alguna comunicación pública *externa* —una manifestación hablada o escrita en lenguaje natural, por ejemplo— depende de la interpretación de las creencias y deseos del que las expresa, de manera que la interpretación de una parte de la maquinaria cognitiva subpersonal interna, tiene que depender en forma inevitable exactamente de la misma cosa: las creencias y deseos de la persona total. El método de Stich para la atribución de contenido depende del mío, y no es un método independiente o alternativo.

Suponga que encontramos en Jones un mecanismo que produce una emisión de "está lloviendo" cada vez que se le pregunta acerca del tema y está lloviendo en la vecindad epistémicamente accesible a Jones. También produce "sí" en repuesta a "¿está lloviendo?" en esas ocasiones. ¿Hemos descubierto la creencia de Jones de que está lloviendo? Es decir, de manera más circunspecta, ¿hemos encontrado el mecanismo que "sirve" a esta creencia en el aparato cognitivo de Jones? Es posible, todo depende de si Jones cree o no que está lloviendo cuando (y solamente cuando) este mecanismo está "en funcionamiento". Quiere decir que tal vez hemos descubierto un mecanismo extraño y absurdo (como el "tumor generador de consentimiento" que imaginé en "Brain Writing and Mind Reading", "*Brainstorms*, pág. 44) que no merece absolutamente ninguna interpretación intencional, o en todo caso no ésta: que es la creencia de que está lloviendo. Necesitamos un estándar frente al cual juzgar nuestros rótulos intencionales para las *illata* de la teoría cognitiva subpersonal. Lo que debemos usar para este estándar es el sistema de

abstracta que fija la creencia y el deseo mediante una especie de proceso hermenéutico que cuenta la historia mejor y más racional que se puede contar. Si descubrimos que Jones pasa los tests correctos —demuestra que en verdad entiende lo que significa la suposición de está lloviendo, por ejemplo— podemos ver confirmada nuestra hipótesis de que hemos dejado al descubierto la realización mecanicista de sus creencias. Pero allí donde encontramos tales deficiencias, proclividades e inactividades tan imperfectas e inadecuadas, disminuiremos, por tanto, nuestros fundamentos para atribuir un contenido de creencia a cualquier mecanismo que encontremos.

He dicho que es improbable que las *illata*, que eventualmente privilegiamos en la psicología académica, se parezcan a las *illata* putativas de la teoría popular lo bastante como para tentarnos a identificarlas. Pero cualesquiera que sean las *illata* que encontremos, las interpretaremos y les asignaremos contenido a la luz de nuestra atribución holística al agente de las creencias y los deseos. Podemos no encontrar en el agente estructuras que se puedan hacer alinear creencia-a-creencia con nuestro catálogo de creencias para el agente de nuestro sistema intencional. Según Stich y Fodor, estaríamos limitados a interpretar este resultado —que todos admiten como posible— como el descubrimiento de que *no existían las creencias después de todo*. La psicología popular era simplemente falsa. A mi criterio, interpretaríamos en cambio este descubrimiento —que es por cierto muy verosímil—, como el descubrimiento de que los sistemas concretos de representación por medio de los cuales el cerebro comprende que los sistemas intencionales simplemente no son de carácter creacional (véase "Más allá de la creencia", capítulo 5).

Por supuesto que a veces tenemos oraciones en la cabeza, lo que no es demasiado sorprendente, si tenemos en cuenta que somos criaturas que usamos un lenguaje. Sin embargo, estas oraciones tienen tanta necesidad de ser interpretadas por la vía de una determinación de nuestras creencias y deseos, como las oraciones que pronunciamos en público. Supongamos que se me ocurren las palabras (solamente "en la cabeza"): "¡Ahora es el momento de la revolución violenta!". ¿*Pensé* entonces ese pensamiento con el contenido de que ahora es el momento de la revolución violenta? Depende, ¿no? ¿De qué? De lo que yo creía, deseaba y quería decir circunstancialmente cuando pronunciaba internamente esas palabras "para mis adentros". Del mismo modo, aun si los "cerebroscopios" demuestran que mientras el chico me daba el cambio acompañaba internamente su transacción con la expresión consciente o inconsciente en su lenguaje natural o en el mentalístico: "Este es el cambio correcto", eso no resolvería la interpretación correcta de esa pizca de lenguaje interno y por tanto no resolvería la interpretación intencional de su acción. Y puesto que él ha cometido un error, no hay ningún catálogo absoluto de sus estados intencionales y acciones en ese momento.

Me mantengo en mi opinión: aun en los casos cotidianos de error que Stich presenta, los problemas de interpretación de creencia que mi punto de vista encontró, *están realmente* en la práctica de la psicología popular, aunque a veces se ocultan detrás de nuestras confabulaciones y excusas. Tampoco se irán debido a la teoría alternativa de atribución de contenido propuesta

por Stich. Esto no quiere decir que tales fenómenos no se puedan describir de manera coherente. Por supuesto, que se los puede describir en forma coherente desde la actitud de diseño o la actitud física, un punto acerca del cual Stich y yo estamos de acuerdo. De manera que no descubro ninguna verdad de la teoría popular de la que lamentablemente tenga que abjurar. Al resistir de esta manera las objeciones de Stich, y mantener a la racionalidad en el basamento de la atribución de creencia y deseo, ¿estoy adoptando lo que Stich llama la "línea dura", o la "línea blanda"? Según Stich la línea dura insiste en que la presunción de racionalidad idealizadora del sistema intencional se encuentra realmente en la práctica popular, de la que deriva la teoría del sistema intencional. La línea blanda "propone cierto manoseo de la noción idealizada de un sistema intencional" para alinearlo más con la práctica popular, la que en verdad (insiste Stich) no apela para nada a las consideraciones de racionalidad. Estas líneas diferentes son invenciones de Stich, nacidas de su frustración ante la tentativa de encontrarles sentido a la expresión de mi punto de vista, que es tanto dura como blanda, es decir, flexible. La *línea flexible* insiste tanto en que la presunción de racionalidad se encuentra en la práctica popular, como en que la racionalidad no es lo que parece ser para algunos teóricos. Así es como la idealización exigirá cierta "frivolidad". ¿Qué digo entonces del ideal de racionalidad explotado conscientemente por el estratega del sistema intencional y como segunda naturaleza por el resto de la gente?

Aquí Stich me encuentra frente a una disyuntiva. Si yo identifico a la racionalidad con la consistencia lógica y la limitación deductiva (y los otros dictados de los sistemas normativos formales tales como la teoría del juego y el cálculo de probabilidades), me desconciertan los absurdos. La limitación deductiva, por ejemplo, es un estado demasiado fuerte, como lo atestigua el caso de Stich de Oscar el ingeniero (véase también Fodor, 1981 a). Si pasando al otro extremo, identifico la racionalidad con *lo que sea que la evolución nos ha proporcionado* o bien caigo en la tautología poco informativa o paso ante ejemplos contrarios evidentes: caso de irracionalidad evolucionada manifiesta. ¿Entonces qué digo que es la racionalidad? No lo digo.

Stich tiene razón. Durante diez años he eludido, insinuado y mantenido afirmaciones que más tarde modifiqué o de las que me retracté. No sabía qué decir y veía problemas dondequiera que me volvía. Con ese *mea culpa* detrás de mí, ahora sin embargo tomaré la ofensiva y daré las que creo que son buenas razones para resistir cautelosamente la exigencia de una declaración acerca de la naturaleza de la racionalidad mientras sigo insistiendo en que una presunción de racionalidad juega el papel crucial que yo le he visto jugar.

Primero, unas pocas palabras acerca de lo que la racionalidad *no* es. No es limitación deductiva. En un pasaje que Stich cita de "Intentional Systems", presentó la sugerencia de que "si S fuera idealmente racional... creería en todas las consecuencias lógicas de todas las creencias (e idealmente, no tendría creencias falsas)" y hago una observación similar en "Los verdaderos creyentes". Ese es después de todo el punto de apoyo lógicamente garantizado de la exigencia universalmente aplicable, la exigencia indefinidamente extensible de que uno crea en las consecuencias "evidentes" de sus

creencias auténticas, totalmente comprendidas. Pero el ejemplo de Stich de Oscar revela gentilmente qué es lo que está mal en permitir que un vínculo ligero expanda las creencias de un agente racional y, como lo demuestra Lawrence Powers en su importante artículo "Knowledge by Deduction" (1978), una teoría de la *adquisición* de conocimientos por deducción tiene mucho trabajo que hacer: uno llega a saber (y creer) lo que todavía no sabía (o creía) deduciendo proposiciones de premisas ya creídas, una idea conocida y "obvia" pero que exige la exposición y defensa muy cuidadosa que le da Powers. Y es importante observar que en el curso de su defensa de lo que podríamos llamar estados cognitivos aislados de las implicaciones, Powers tiene que recurrir al neologismo y la advertencia: debemos hablar acerca de lo que nuestro agente "seudocree" y "seudosabe" (pág. 360 y sigs.). En realidad, le recuerda a uno el propio neologismo útil de Stich para los estados tipo creencia que carecen de la fecundidad lógica de las creencias: "estados subdoxásticos" (Stich, 1978b).

Tampoco es la racionalidad una consistencia lógica perfecta, aunque el *descubrimiento* de contradicción entre proposiciones con las que uno se siente inclinado a asentir siempre es, por supuesto, una ocasión para hacer sonar la alarma epistémica (de Sousa, 1971). La inconsistencia, cuando se la descubre, debe ser eliminada de una forma u otra, por supuesto, pero hacer que desenterrar la inconsistencia sea la meta predominante del conocedor, llevaría a hundir el sistema cognitivo en operaciones contables y de búsqueda con exclusión de todos los demás aspectos de la actividad (Cherniak, 1986 y Darmstadter, 1971). Ahora bien, ¿cómo puedo hablar de este modo acerca de mi inconsistencia, dada mi explicación de las condiciones de la atribución correcta de creencias? ¿Quién dijo algo acerca de la inconsistencia de las *creencias*? Cuando uno ingresa al dominio de las consideraciones acerca del sabio diseño de las estructuras y operaciones cognitivas, ya ha dejado atrás la creencia propiamente dicha y está discutiendo, en realidad, características estructuralmente identificadas con rótulos intencionales más o menos aptos (véanse "Tres clases de psicología intencional" y *Brainstorms*, págs. 26-27).

Si así no identifico a la racionalidad con la consistencia y la limitación deductiva, ¿cuál podría ser entonces mi estandarte? Si me vuelvo hacia las consideraciones evolucionistas, sugiere Stich, "teorías tan establecidas como la lógica deductiva e inductiva, la teoría de la decisión y la teoría del juego no serán de ninguna ayuda para evaluar qué 'debería creer' un organismo". Esto no es cierto. El teórico que renuncia a la afirmación de que estos formalismos son el hito final de la racionalidad todavía puede volverse a ellos en busca de ayuda, todavía puede explotarlos mientras critica (sobre la base de la irracionalidad) y reformula estrategias, diseños, interpretaciones. La analogía es imperfecta pero del mismo modo en que se puede buscar la ayuda de un buen diccionario o un buen libro de gramática para apoyar la crítica de la ortografía, la elección de palabras o la gramática de alguien, se puede apelar a la autoridad revocable de, digamos, la teoría de la decisión para objetar la formulación estratégica de alguien. También se puede rechazar como incorrecto —o irracional— el consejo que uno recibe de un diccionario, una gramática, una lógica o cualquier otra teoría normativa, por bien afirmada que esté (Cohen, 1981; Stich y Nisbett, 1980).

¿Y qué hay de las consideraciones evolucionistas? Tengo cuidado en *no* definir la racionalidad en los términos que la evolución nos ha proporcionado, así es que evito la tautología franca. No obstante, la relación, que yo afirmo, se mantiene entre la racionalidad y la evolución es más poderosa que lo que Stich concederá. Como él observa, yo afirmo que si un organismo es el producto de la selección natural podemos dar por sentado que la mayor parte de sus creencias serán ciertas y que la mayoría de sus estrategias de formación de creencias serán racionales. Stich no está de acuerdo: "Sencillamente no es verdad que la selección natural favorezca las creencias verdaderas sobre las falsas", puesto que todo lo que la selección natural privilegia son las creencias "que producen ventajas selectivas" y "hay muchas circunstancias ambientales en las que las creencias falsas serán más útiles que las verdaderas". Yo no creo que sea evidente que pueda ser ventajoso alguna vez estar diseñado para llegar a creencias falsas acerca del mundo, pero he alegado que hay circunstancias descriptibles —circunstancias raras— donde eso puede suceder, de manera que concuerdo con Stich en este punto: "*Es mejor estar a salvo que tener que lamentarse* es una política que se recomienda a la selección natural", dice Stich, haciéndose eco de mi afirmación en "Tres clases de psicología intencional". Equivocarse con prudencia es una buena estrategia bien reconocida, por lo tanto se puede esperar que la naturaleza la haya valorado en las ocasiones en que apareció.

¿Pero llega esto de algún modo a impugnar mi afirmación de que la selección natural garantiza que la mayoría de las creencias de un organismo sean verdad, la mayoría de sus estrategias racionales? Creo que no. Más aun, incluso cuando una estrategia sea como concedemos que muy bien pueda ser, una estrategia "visiblemente nula" que funciona la mayor parte del tiempo en los contextos en los que se la invoca, ¿prueba esto que sea una estrategia irracional? Sólo si uno todavía se aferra a los ideales de la Intro Lógica para su modelo de racionalidad. No se trata ni siquiera de que no hayan cánones de racionalidad académica reconocidos en oposición a los de los especialistas en Lógica a los que uno podría recurrir. Herbert Simon es justamente famoso por sostener que *satisfacer es racional* en muchos casos: por ejemplo, lanzarse a conclusiones posiblemene "sin valor" cuando el coste de efectuar mayores cálculos probablemente excede el coste de obtener la respuesta equivocada. Creo que tiene razón, de manera que por mi parte yo no ataría a la racionalidad a cánones que prohibieran esas prácticas. Stich declara:

> Siempre que reconozcamos una diferencia entre una teoría normativa de inferencia o toma de decisiones y un conjunto de prácticas deductivas que (en el entorno correcto) generalmente consiguen la respuesta correcta (o selectivamente útil), quedará claro que las dos no tienen que coincidir y generalmente no coinciden (págs. 53-54).

Esta es una afirmación sorprendente, puesto que hay teorías normativas para distintos propósitos, incluyendo los de "conseguir generalmente la respuesta correcta". Si uno considera a éstos como en desacuerdo unos con los otros, comete un error. Se podría sostener la lógica deductiva para recomen-

dar que frente a la incertidumbre o la falta de información uno debería simplemente *no moverse ni inferir nada*, un mal consejo para un ser en un mundo atareado pero un consejo muy bueno si la meta es evitar la falsedad a cualquier precio. Es mejor reconocer los distintos usos a los que se pueden aplicar esas estrategias y dejar que la racionalidad consista en parte en un buen sentido de cuándo confiar en qué. (También es útil que nos recordemos a nosotros mismos que sólo una fracción diminuta de todos los "animales racionales" que alguna vez vivieron se valió conscientemente de alguna técnica formal de las teorías normativas que se han propuesto.)

No hay duda de que el concepto de la racionalidad es escurridizo. Parece que estamos de acuerdo en que un sistema sería llamado irracional de manera inadecuada si, aunque su desempeño *diseñado*, *normal* fuera impecable (según los estándares de las normas pertinentes), tuviera un *funcionamiento defectuoso* ocasional. Pero, por supuesto, que un sistema particularmente propenso al mal funcionamiento que no es corregido, no sería precisamente un sistema bien diseñado. En este sentido, un sistema que fuera infalible y capaz de protegerse sería mejor. ¿Pero cuál sería mejor —cuál sería más racional— considerando todo: un sistema muy lento pero virtualmente autoprotector, o uno muy veloz pero libre sólo en un 90 % del mal funcionamiento? Depende de la aplicación y hasta hay cánones normativos para evaluar esas elecciones en algunas circunstancias.

Quiero usar "racional" como un término de uso general de aprobación cognitiva, que requiere mantener sólo lealtades condicionales y enmendables entre la racionalidad así considerada y los métodos propuestos (o hasta universalmente aclamados) para avanzar, cognitivamente, en el mundo. Tengo entendido que este uso de los términos es completamente corriente, y apelo a la racionalidad de quienes proponen disciplinas o prácticas cognitivas para requerir esta comprensión de la noción. ¿Por ejemplo, a qué estarían apelando Anderson y Belnap (1974), qué podrían estar suponiendo acerca de su público cuando recomiendan su explicación del vínculo por encima de sus rivales, si no a una racionalidad presumiblemente compartida tal que es una *pregunta abierta* acerca de cuál es el sistema formal que mejor lo capta? O considere este comentario acerca del descubrimiento de que una memoria compartimentalizada es una condición necesaria para la cognición efectiva en un mundo complejo, presionado por el tiempo:

> Ahora podemos apreciar tanto los costes como los beneficios de esta estrategia; *prima facie*, la conducta resultante se puede caracterizar como desviaciones de la racionalidad, pero, basándose en la presunción de que la búsqueda exhaustiva de la memoria no es factible, tal organización de la memoria es totalmente aconsejable a pesar de sus costes. De igual forma, la acción de una persona puede parecer irracional cuando se la considera aisladamente, pero puede ser racional cuando se la considera más ampliamente como parte del precio que vale la pena pagar por un buen manejo de la memoria. (Cherniak, 1983, pág. 23).

La aseveración es que es racional ser inconsistente a veces, no la aseveración seudoparadójica de que es racional ser irracional a veces. Como lo demuestra el ejemplo, el concepto de racionalidad es sistemáticamente preteórico. Uno puede, entonces, rehusarse a *identificar* la racionalidad con las

características de algún sistema formal o el resultado de algún proceso y, no obstante, hacer apelaciones al concepto y a las afirmaciones acerca de las apelaciones a él (tal como la mía) sin por ello eludir una obligación de precisión.

Cuando nos apoyamos en nuestro concepto preteórico de racionalidad, confiamos en nuestras intuiciones compartidas —cuando *son* compartidas, por supuesto— acerca de lo que tiene sentido. ¿En qué otra cosa, finalmente, se podría confiar? Cuando consideramos lo que *deberíamos hacer*, nuestras reflexiones nos conducen eventualmente a una consideración de lo que ***en realidad hacemos;*** esto es ineludible, puesto que un catálogo de nuestros juicios considerados intuitivos acerca de lo que tendríamos que hacer es tanto un compendio de lo que realmente pensamos, como un ejemplo reluciente (a nuestra luz, ¿qué otra cosa?) de cómo tendríamos que pensar:

> De este modo, qué y cómo realmente pensamos es una evidencia en favor de los principios de racionalidad acerca de qué y cómo tendríamos que pensar. Este es en sí mismo un principio metodológico de racionalidad; llamémoslo el *Principio Factunorm*. Estamos aceptando (implícitamente) el Principio Factunorm cada vez que tratamos de determinar qué y cómo tendríamos que pensar. Puesto qué y cómo pensamos allí es correcto y es, por lo tanto, una evidencia en favor de qué y cómo tendríamos que pensar, no podemos determinar qué y cómo tendríamos que pensar. (Wertheimer, 1974, págs. 110-111).

Ahora parecerá que estoy retrocediendo al propio punto de vista de Stich, el punto de vista de que cuando atribuimos creencias y otros estados intencionales a otros, lo hacemos comparándolos con nosotros, al proyectarnos a nosotros mismos dentro de sus estados de ánimo. Uno no pregunta "¿qué tendría que creer este ser?" sino "¿qué creería yo si estuviera en su lugar?". (Le he sugerido a Stich que llame a su punto de vista *solipsismo ideológico*, pero aparentemente él siente que esto crearía confusión con alguna otra doctrina.) Stich contrasta su punto de vista con el mío y asevera que la noción de racionalidad idealizada *no juega absolutamente ningún papel* (el énfasis es de Stich) en su explicación. "Al atribuir contenido a los estados de creencia, medimos a los demás no según un estándar idealizado, sino según nosotros mismos. "Pero por las razones que acabamos de dar, medir según nosotros mismos" *es* medir según un estándar idealizado.

En algún punto Stich observa que "puesto que nos consideramos a nosotros mismos como aproximados a la racionalidad, esto explica el hecho, notado por Dennett, de que la descripción intencional vacila cuando está frente a la irracionalidad flagrante. "El debe admitir, entonces, que puesto que nos tomamos a nosotros mismos como próximos a la racionalidad, es también cierto que los resultados de mi método y su método coincidirán muy de cerca. El, al preguntar... "¿qué haría yo si...?" y yo, al preguntar "¿qué tendría él que hacer...?" llegamos típicamente a la misma explicación, puesto que Stich supondrá típicamente que lo que tendría que hacer es lo que yo haría si estuviera en su lugar. Si los métodos fueran verdaderamente equivalentes en lo extensivo, uno podría muy bien preguntarse acerca del punto en disputa,

pero, ¿no hay lugar para que dos métodos difieran en casos especiales? Veamos.

¿No podría ser así? Stich, conocedor de su lamentable y embarazosa tendencia a afirmar lo consecuente, les imputa esta misma tendencia a aquellos cuyas creencias y deseos está tratando de desentrañar. Hace esto en lugar de suponer que podrían estar libres de su propia flaqueza particular, pero que podrían ser culpables de otras. Una historia poco probable. He aquí una mejor. Habiendo aprendido algo acerca de la "disonancia cognitiva", Stich está ahora preparado para encontrar, tanto en sí mismo como en otros, la resolución de la disonancia cognitiva el hecho de favorecer una creencia autojustificativa por encima de una creencia menos cómoda sustentada por la evidencia. Este es un muy buen ejemplo de la clase de descubrimiento empírico que se puede usar para afinar la actitud intencional, sugiriendo hipótesis a ser probadas por el que las atribuye, pero, ¿cómo diría Stich que tendría algo que ver con *nosotros mismos*, y cómo se pondría en uso efectivo este descubrimiento, independientemente de la presunción idealizadora? Primero, ¿no va a ser una pregunta empírica si toda la gente responde a la disonancia cognitiva como lo hacemos nosotros? Si Stich arma esta (aparente) proclividad subóptima en su propio método de atribución, él se priva de la posibilidad de descubrir variedades de creyentes felizmente inmunes a esta patología.

Más aun, consideramos cómo esa presunción de suboptimidad sería utilizada en un caso real. Jones acaba de pasar tres meses de trabajo duro construyendo una ampliación de su casa; parece horrible. Hay que hacer algo para resolver la incómoda disonancia cognitiva. Cuente con Jones para deslizarse hacia alguna creencia que salvará la situación. ¿Pero cuál? El podría llegar a creer que el objetivo del proyecto era, en realidad, aprenderlo todo acerca de la carpintería mediante el expediente realmente poco costoso de construir un anexo barato. O podría llegar a creer que la arremetida audaz del anexo no es más que el toque que distingue su casa, por otra parte vulgar si bien "de buen gusto" del conjunto de las casas del barrio. O... muchas variaciones posibles. Pero, ¿cuál de ellas es lo que realmente cree que se decidirá observando lo que dice y hace, y preguntando luego: ¿qué creencias y deseos volverían racionales esos actos? Y, sea cual fuere la ilusión que se adopte, debe estar —y estará— cuidadosamente rodeada por un material de apoyo plausible, generable sobre la presunción poco realista que la ilusión es una creencia completamente racional. Dado que ya conocemos a Jones, tendríamos que poder predecir qué ilusión alentadora sería la más atractiva y eficaz para él, es decir, cuál sería más coherente con el resto de la trama de sus creencias. De manera que, aun en caso de una disonancia cognitiva, cuando las creencias que atribuimos no son óptimas a criterio de nadie, el test de coherencia racional es la medida preponderante de nuestras atribuciones.

No veo cómo se puede probar que mi método y el de Stich producen resultados diferentes, pero tampoco veo que no se pueda. No tengo bastante claro qué es exactamente lo que Stich afirma. Una idea interesante que acecha en el punto de vista de Stich es que cuando interpretamos a otros no lo hacemos tanto *teorizando* acerca de ellos sino como usándonos a nosotros mismos *como ordenadores análogos* que producen un resultado. Al querer saber más acerca de su estado de ánimo, de algún modo me pongo en su lu-

gar, o lo más cerca de él que pueda llegar y veo lo que por consiguiente pienso (quiero, hago...). Hay mucho de enigmático en semejante idea. ¿Cómo puede funcionar sin ser una clase de teorización al fin? Porque el estado en el que me pongo no es creencia sino creencia fingida. Si finjo que soy un puente colgante y me pregunto qué voy a hacer cuando sople el viento, "lo que se me ocurre" en mi estado fingido de cuán elaborado sea mi conocimiento de la física y la ingeniería de los puentes colgantes. ¿Por qué debería ser diferente mi simulación de sus creencias? En ambos casos, es necesario el conocimiento del objeto imitado para impulsar la "simulación" fingida, y el conocimiento debe estar organizado como algo parecido a una teoría. Más aun, establecer que realmente llegamos de algún modo a nuestras interpretaciones de otros por medio de algo parecido a la simulación y la autoobservación, no demostraría por sí mismo que la pregunta guía de nuestros esfuerzos es "¿qué creería yo?" *como contraria a* "¿qué tendría que creer él?". Un atribuidor prudente podría mostrar la diferencia mediante el truco de la empatía o el fingimiento para *generar* un conjunto de atribuciones candidatas a *ser probadas* contra su "teoría" del otro, antes de decidirse por ellas. Observe que el tema está lejos de ser claro aun en el caso de la *auto*atribución imaginada. ¿Cuál sería su estado de ánimo si le dijeran que tiene tres semanas de vida? ¿Qué piensa de esto? Probablemente varias cosas; usted simula un poco y ve qué *diría, pensaría* y demás, y también reflexiona acerca de qué clase de persona usted cree ser, de manera de poder llegar a la conclusión de que una persona *así* creería —tendría que creer— o desear esto o aquello.

Cierro con una réplica final. Stich trata de avergonzarme al terminar con una serie de preguntas retóricas acerca de lo que *tendría que creer* una rana, puesto que yo he decidido que lo que una rana *verdaderamente* cree depende de esas preguntas. Admito que es problemático responder a esas preguntas aun en las mejores condiciones, pero no considero que eso sea embarazoso. Contesto con una pregunta retórica propia: ¿supone Stich que el contenido exacto de lo que una rana realmente cree sea más probable de determinar?

Reflexiones: Cuando las ranas (y otros) cometen errores

Stich tiene un don muy bueno para formular las preguntas correctas, no las odiosas preguntas matadoras sino las significativamente provocadoras de pensar mejor en algo. Ahora, con más tiempo para pensar acerca de sus desafíos y más espacio en el que responder a ellos, quiero ocuparme con más detalle de dos de ellos: el vendedor de limonada y la rana. ¿Qué creen en realidad? Mis respuestas a estas dos preguntas llevan a ulteriores reflexiones acerca de por qué el Realismo acerca de las creencias es tan poco realista.

El error del vendedor de limonada

El vendedor de limonada ha cometido un error simple y Stich extrae correctamente la llamativa implicación de mi punto de vista acerca de esos

errores: debo afirmar que a menos que el error tenga alguna de las etiologías normales, que podríamos llamar periféricas —él ha observado mal las monedas que tiene en la mano o ha estado mal informado acerca de su valor o de las verdades de la aritmética: debe haber un punto irracional, una brecha que no se puede interpretar, en la historia que contamos acerca de él desde la actitud intencional. Dicho en forma contundente, a mi criterio ¡no hay manera de decir lo que alguien verdaderamente cree cada vez que comete un error cognitivo! Stich sugiere que esto es evidentemente falso. Presumiblemente todos esos errores son algún error particular, de donde podría deducirse que siempre habrá (aun cuando no podamos encontrarla) una historia de creencia completa y mejor para contar acerca de él.

Una historia completa de creencia le asignaría veracidad o falsedad a toda atribución de creencia pertinente. Piénselo sobre el modelo de un interrogatorio en un juicio. ¿Creyó el acusado que daba el vuelto correcto? ¿Creyó que las dos monedas que entregó eran de diez y de un centavo? ¿Conocía o no el valor de un penique? ¿Sí o no? Admito que realmente parece obvio que todas éstas son preguntas justas que tienen que tener respuestas correctas aun cuando nadie pueda determinarlas, pero afirmo que ésta es una ilusión. Es pariente cercano de la ilusión, señalada primero por Quine, de que tiene que haber un manual de traducción mejor entre dos idiomas aunque no podamos encontrarlo. Quine ha sido desafiado con frecuencia a dar un ejemplo convincente de un caso auténtico (aunque sea imaginario) de la incertidumbre en la traducción radical (muy memorablemente en la conferencia sobre Intencionalidad, Lenguaje y Traducción que aparece en forma de antología en *Synthese*, 1974). No ha podido dar un ejemplo detallado y realista, no porque su tesis sea falsa sino porque postula la posibilidad de un caso que está en equilibrio en el filo de una navaja: dos manuales de traducción radicalmente en desacuerdo de manera tal que *toda* la información pertinente disponible no logra favorecer a uno por encima del otro. El mundo real aborrece los filos de las navajas todavía más que los vacíos. Casi inevitablemente los detalles desprolijos empiezan a amontonarse más rápido de un lado que de otro arruinando cualquier punto de vista realista para el papel de Ejemplo de Quine.

Se podría anticipar un destino semejante para mi afirmación de incertidumbre y, no obstante, me parece más fácil mantener la afirmación de que la brecha es real e imposible de cerrar en el caso de los errores cognitivos (véase Wheeler, 1986). Puesto que, aunque en todos los casos que puedo imaginar, los detalles no tan desprolijos verdaderamente aumentan a favor de una u otra historia acerca de la naturaleza del error, ellos no arrojan ninguna luz indiscutible acerca de *lo que la persona creía*. Tenga en cuenta cómo podríamos terminar la historia del vendedor de limonada.

Supongamos que pruebas concluyentes en los laboratorios de científicos cognitivos determinaron que una vez que ha "entrado en calor", él da el cambio de manera semiautomática mientras piensa en otras cosas. Un "módulo" temporal dador de cambios está alojado en su cerebro y la imagen por resonancia magnética nuclear demuestra que funcionaba en ese momento más aun se descubre que 350 milésimos de segundo antes de que el módulo dador de cambio terminara su tarea, la sonrisa de una muchacha que iba en

un auto que pasó, atrajo su atención, haciendo que el submódulo que tomaba los peniques hiciera mutis después de recoger un solo penique en lugar de tres. *Las realimentaciones* visuales y táctiles, que normalmente hubieran detectado el error, fueron recibidas y pasadas por alto debido a un cambio de atención provocado por mi ruidosa manera de sorber la limonada y el hecho de que el módulo informó acerca de la terminación con éxito de su tarea antes de cerrar.

Ahora bien, ¿qué creía él? ¿Creyó que me dio una moneda de diez centavos y una de un centavo de cambio? El veía lo que estaba haciendo y no sufría ninguna alucinación visual (*ex hypothesi* en este relato). Si se lo hubiera pedido inmediatamente después de que yo tomé el cambio que dijera qué me había dado, él habría dicho una moneda de diez y una de uno. De todos los poderes que se podría esperar que tal creencia tuviera, el único que falta en este caso, podemos suponer, es el poder de *iniciar* una corrección (recuerde que en cuanto se le señala el hecho de que me dio una moneda de diez y una de uno, reconoce su error y lo corrige).

"Tenía la creencia, pero simplemente no le prestó atención."

"Muy bien; ¿creyó también que me daba trece centavos de cambio?

"Sí, pero no le prestó atención a esa creencia tampoco. Si les hubiera prestado atención a ambas habría notado la contradicción.

"¿No pudo haberles prestado atención a ambas y *no* haber reconocido la contradicción?

"No; si él *entiende en verdad* esas dos proposiciones *debe* reconocer su incompatibilidad."

¿Pero no es igualmente constitutivo de la creencia en tal proposición que uno se da cuenta cuando es contradecido? El no debe *creer verdaderamente* esas proposiciones, puesto que entre sus poderes definitivos como creencias debe estar el de que causan alarma, atraen la atención hacia ellas mismas cuando aceptan o alientan proposiciones contradictorias.

Este debate podría seguir, pero observe cómo ha dejado atrás la psicología cognitiva subpersonal imaginada. Esto se debe a que los mecanismos imaginados tienen tanta necesidad de interpretación como la conducta exterior, y las mismas reglas sirven; de ahí el conflicto. Los fundamentos para decir que "le di una moneda de diez y una de uno", representados tácitamente por el estado del módulo más la memoria visual (etc.) son poderosos, pero también lo son los fundamentos para decir que no son lo bastante poderosos como para una creencia plena. Sostengo que lo mismo ocurrirá con cualquier relato. Siempre se puede resolver despejando los casos indefinidos hacia un lado u otro a los fines de la prolijidad. Pero entonces no se están usando los hechos acerca de los procesos internos como indicios que llevan a descubrimientos acerca de las creencias de alguien, sino que se los está usando simplemente como fundamentos más conductistas para las atribuciones del tipo ya recogido y perdiendo luego la paciencia. Tal como: "¡Pero así como Bill le *dijo* que *p*!" es un fundamento poderoso pero revocable para atribuirle la creencia de que *p* (él puede no haber entendido y puede haberse olvidado), así es "Pero el módulo le *dijo* al resto del sistema que había dado bien el cambio".

Esto sugiere que estamos mirando al lugar equivocado cuando miramos los módulos "periféricos", los órganos y efectuadores de los sentidos. ¿No tendríamos que "centrar" más la atención en la arena donde ocurre la "fijación de la creencia"? (Fodor, 1983: Dennett, 1984a). El mito de que se podría encontrar la verdad escrita en pequeños compartimientos en la Caja de Creencia tarda en desaparecer. (Véanse los dos capítulos siguientes.) Sin embargo, es menos compulsivo en el caso de objetos no humanos de la actitud intencional, tales como las ranas.

Psicología de la rana

Stich cierra su trabajo con una serie de preguntas:

> ¿Debería creer la rana que hay un insecto volando hacia la derecha? ¿O simplemente que allí hay algo para comer? ¿O tal vez tendría que tener sólo una creencia condicional: si chasquea la lengua de cierta manera algo delicioso terminará en su boca? Supongamos que la mosca es de una especie que les causa indigestión aguda a las ranas. ¿Tendría la rana que creer esto? ¿Constituye alguna diferencia a cuántas compañeras ranas ha visto sufrir después de mascar bichos parecidos? (págs. 60-61).

Miremos más de cerca el contraste sugerido entre nosotros y las ranas. La rana está ubicada en su ambiente de un modo muy complicado, bañada de información potencialmente útil gracias a las miriadas de interacciones entre sus receptores sensoriales y los ítems del mundo que la rodea. Es capaz, en cualquier momento de su vigilia, de explotar esa ola de información en formas que se pueden resumir aproximadamente diciendo cosas como ésta:

> Ahora la rana ve que su sombra aparece en forma amenazadora. Quiere escapar de usted, cree que usted está detrás de ella y puesto que no ve su red haciendo guardia en la abertura de la izquierda, cree que es por allí que debe escapar, por tanto saltará hacia la izquierda.

Consideremos de uno en uno los modismos mentalistas que se explotan en la adopción de la actitud intencional hacia la rana. Las ranas tienen ojos, de manera que, por supuesto, en algún sentido, *ven*. ¿Ven verdaderamente? (¿Ve verdaderamente una pulga? ¿Ve verdaderamente un molusco de ojos azules? Tiene docenas de ojos que le permiten reaccionar ante modelos movibles de luz y sombra.) Ahora sabemos que la vista de la rana es muy distinta de la nuestra. Si nuestros antepasados hubieran supuesto cuán pobre era la vista de la rana comparada con la nuestra, podrían no haber dicho que las ranas ven verdaderamente. ¿Cuenta como visión *cualquier* beneficio guía de acciones derivado de la fotosensibilidad? ¿Dónde trazamos la línea? El problema es menor y mayormente sólo léxico, pero se puede convertir en un acertijo filosófico si se elige exactamente el tipo correcto de caso fronterizo.

¿Y qué hay de los no videntes a los que se les han colocado prótesis visuales? Usan una simple cámara de televisión sobre la cabeza y su señal se extiende por un conjunto de varios cientos de cosquilleos en la espalda o el

vientre. Esta gente puede entrenarse para responder perceptivamente a los imperfectos modelos de luz y oscuridad detectables por sus cámaras de televisión al punto de poder identificar letras del alfabeto y de ahí "leer" grandes letreros. ¿Esa prótesis les permite *ver*? El fenómeno es fascinante, pero el acertijo filosófico no lo es. Una vez que uno sabe exactamente qué poderes perceptivos disfrutados por la gente de vista normal pueden (y no pueden) adquirirse con ese dispositivo, lo único que queda es tomar una decisión léxica táctica acerca de si sería o no engañoso llamar visión a ese fenómeno.

Desde la perspectiva de la primera persona esto parece excluir el punto más importante: *cómo es* recibir información acerca del mundo distal de esa manera. Pero en realidad esto no ha sido excluido. Al estudiar los poderes perceptivos de los usuarios adeptos del dispositivo, uno se entera que los hormigueos sobre la piel pronto pasan inadvertidos; "su punto de vista" se traslada a un punto encima de sus cabezas y oscilan cuando las giran (Livingston, 1978). ¿No determina esto que ésta es una clase de visión? No para algunos filósofos. Se preocupan por si semejante sistema le proporcionaría a su usuario lo que ellos consideran ser las propiedades esenciales intrínsecas de la visión real. Pero cualquiera que sea la propiedad ulterior esencial intrínseca de la visión que ellos se ocupen de exudar, procurando nuestra aprobación, deben admitir que su suposición de que las ranas la tienen (o no la tienen) es pura conjetura. O no tenemos ninguna forma de saber si las ranas verdaderamente ven o podemos atenernos a la capacidad de sus sistemas de reunir información, en cuyo caso, la visión humana protésica es, obviamente, una manera de ver igual que las demás. [Como de costumbre, el punto de vista del tercero progresa, mientras que el del primero se agota en una pregunta sistemáticamente misteriosa acerca de propiedades intrínsecas imaginadas (véase *Brainstorms*, capítulo 11, y "Qining Qualia", por aparecer d).]

De manera que la rana ve. Las ranas muestran también la clase de conducta evasiva tramposa que nos lleva a interpretarla como *querer escapar*, una interpretación que adquiere mayor sentido porque nos es muy fácil pensar en buenas y justificadas razones para que las ranas "quieran" mantenerse a buena distancia de nosotros. ¿Y cómo deberíamos llamar a la contribución de esos ojos al control de esas extremidades si no *creencias* acerca de la posición de esto o aquello? ¿Pero tienen realmente las ranas creencias y deseos? La brecha entre nosotros y las ranas parece más grande en esto que en la vista. Ninguna rana creería que las ballenas son mamíferos, o que el viernes viene después del jueves, o querría una pizza o tendría esperanzas de visitar Río. Tampoco tiene que ver con lo lejos que están esos temas de los intereses de las ranas. Creo que las ranas tienen patas enmarañadas y que atrapan los insectos voladores con la lengua, pero (parece que) no se podría afirmar debidamente que las ranas creen esas proposiciones. ¿*Quiere* alguna rana algo tanto como *encontrar un montón de insectos hoy*?

Aun cuando, en el mejor de los casos posibles, nos sintamos cómodos atribuyéndole una creencia a la rana —tal vez la creencia de que un depredador grande que está detrás de ella va a atacar— en apariencia no se dispone de los principios necesarios para volver preciso el contenido de una creencia atribuida. Eso es lo que causa impresión a Stich. ¿Qué concepto de un

depredador tiene la rana? ¿Podemos distinguir entre su creencia de que un depredador está detrás de ella y la de que (más vagamente) algo que hay que evitar está detrás de ella? Cuando anda en busca de moscas, ¿se puede decir que busca moscas *qua* moscas, o simplemente *qua* cosas comestibles oscuras que pasan volando rápidamente o *qua* algo todavía menos específico (véase Dennett, 1969, capítulos 4 y 10)? Las oportunidades que se prestan para la caracterización de las ranas en términos de sus "creencias" y "deseos" son ridículamente reducidas e imprecisas en comparación con las nuestras. Davidson (1975, pág. 16) y Dretske (1985, pág. 30) han afirmado algo parecido.

Y, no obstante, este modo antropomorfizante de organizar y simplificar nuestras expectativas acerca de los movimientos siguientes de la rana es compulsiva y útil. Tratar a las ranas, pájaros, monos, delfines, langostas, abejas —y no sólo a hombres, mujeres y niños— desde la actitud intencional, no sólo surge naturalmente, sino que también funciona muy bien dentro de su reducido radio de acción. Trate de cazar ranas sin ella.

La amplia diferencia entre el alcance de un creyente humano adulto y una rana sugiere que la aplicación de la discusión de la creencia y la del deseo a las ranas no es más que una extensión metafórica de su uso apropiado, aplicado a los seres humanos, los verdaderos creyentes. Esta sugerencia es inmensamente persuasiva. Es probable que sea la única fuente más poderosa de escepticismo hacia mi posición, que sostiene que no hay nada más difícil en el hecho de que *nosotros* tengamos creencias y deseos que el que seamos muy fáciles de predecir (como la rana, pero más) desde la actitud intencional.

Los críticos me *atacan* desde dos lados: aquellos que creen que sólo nosotros, los humanos, tenemos creencias y aquellos que piensan que, en realidad, no hay tales creencias. Puedo refutar a los oponentes a la vez demostrando exactamente dónde, al rastrear nuestro terreno compartido, ellos se apartan de mí y diagnostican después el sutil error que yo supongo que ellos cometen.

Volvamos a tomar en cuenta a la rana. Cuando está inmóvil en la hoja de nenúfar, su sistema nervioso está en plena intrincada actividad. Los productos de millones de interacciones entre los fotones, ondas de presión acústica, células receptoras, secreciones internas, y demás, interactúan unas con las otras para producir todavía más actividades, que eventualmente producen entre la suma de todas, los pulsos eferentes que contraen los músculos de la pata de la rana y la lanzan de un salto hacia la izquierda dentro de la red. Uno podría haber pronosticado ese salto si hubiera sabido bastante biología y calculado las interacciones a partir de una copia exacta del sistema nervioso de la rana. Esa sería una predicción desde la actitud de diseño. En principio, uno podría haber ignorado los principios biológicos pero haber sabido bastante física como para predecir el salto de la rana a partir de un cálculo abundante de las interacciones energéticas de todas las partes, desde la actitud física.

El famoso demonio de Laplace no tiene que tener el concepto de una neurona eferente cuya *función* es transportar una señal que causa una contracción del músculo de la pata de la rana; puede predecir el salto con sólo rastrear los efectos físicos esperados de todas aquellas excursiones de la

membrana iónica por el sendero que el biólogo identificaría, desde la actitud de diseño, como el neuroeje.

En principio, entonces, la conducta de la rana se puede calcular y explicar sin ninguna invocación a la "psicología", ni desde la actitud del piso bajo de la física o la actitud de diseño ligeramente elevada de la biología. La *psicología de la rana* se puede considerar como un atajo prácticamente útil pero teóricamente injustificado: algunos cálculos primitivos pragmáticos para simplificar en exceso la complejidad. La psicología de la rana es gratuita sólo en este sentido: las predicciones desde la actitud física o la biología tienen hegemonía sobre las de la actitud intencional; ninguna imprevisible "propiedad emergente" intermedia o "efecto crítico masivo" es un obstáculo, ni amenaza con falsear las trabajosas predicciones desde las actitudes más bajas; utópicamente la complejidad es (prácticamente) un obstáculo para la predicción.

El demonio de Laplace diría exactamente lo mismo acerca de la biología de la rana, por supuesto; todo lo que en verdad hace falta es la física de la rana, diría. Pero aun si admitimos que el demonio tiene razón en principio, parece, sin duda, que las categorías de la biología de la rana tienen una realidad más robusta que las categorías de la psicología de la rana. Después de todo, el sistema nervioso de la rana se puede ver. Se pueden individualizar las neuronas bajo un microscopio de alto poder y medir el estado en que está una neurona con un electrodo implantado.

He aquí lo que muchos, quizá la mayoría de quienes escriben sobre este tema dirían: es obvio que hablar del *deseo de escapar de la rana* o de *la creencia de que usted está detrás de ella* no es precisamente hablar de un estado especial de su sistema nervioso, sino sólo aludir, de manera indirecta e imprecisa, a cierta tendencia dispositiva o recurrir a su estado neurofisiológico corriente y dócil, en principio, a una descripción de grano fino exacta y exhaustiva. Es una llana y eficiente *façon de parler* pero hay otras maneras más científicas de hablar. Eventualmente encontraremos la correcta para describir el desempeño del sistema nervioso de la rana. Esta teoría de la neurofisiología de la rana puede muy bien describir las cosas en los términos de información llevada de aquí para allá, procesada así y asá, por varios subcomponentes neuronales pero guardará silencio acerca de las creencias y los deseos de la rana porque, estrictamente hablando, la rana no tiene ninguno.

Todo esto es muy plausible y por cierto que yo lo apruebo todo, excepto por el tono y la última línea: "estrictamente hablando" implica un contraste equivocado. Le sugiere a una escuela de mis críticos (Fodor, Dretske y otros "realistas" acerca de la creencia) que somos diferentes: nosotros, los seres humanos, tenemos verdaderamente creencias y deseos, y cuando le atribuimos una creencia a un ser humano, esto es una declaración no metafórica capaz de una precisión considerable. No tengo del todo claro dónde trazan la línea los realistas individuales. Fodor (1986) está seguro de que los paramecios no tienen creencias pero quizá las ranas logren ingresar en el círculo encantado.

La otra escuela de críticos (Stich, los Churchland y otros "eliminacionistas" de la creencia) aceptan el contraste implícito y con él la opinión del primer grupo de lo que sería *realmente tener una creencia*. Estrictamente

hablando, coinciden, las ranas no tienen creencias pero estrictamente hablando, ¡nosotros tampoco! No existen tales creencias.

Coincido en que si las creencias tuvieran que ser lo que los realistas creen que son, no habría creencias ni para las ranas ni para ninguno de nosotros. No me siento tentado, como ambos grupos de críticos lo están, por el contraste entre la rana, tal como se la describió, y nosotros. Por supuesto que nadie supone que somos completamente distintos a la rana. Nosotros también estamos inundados de información y tenemos sistemas nerviosos astronómicamente complicados, y muchas de las transacciones que ocurren dentro de ellos muestran un carácter específico escaso y propio de la rana cuando se la caracteriza desde la actitud intencional. Por ejemplo, ¿cuál es exactamente el contenido de su "creencia" perceptiva cuando una sombra amenazadora en su campo visual lo hace echarse atrás?

Las ilusiones del realismo

Sin embargo, a los realistas les parece que además de esos rasgos de nuestra propia conducta que admiten una caracterización intencional meramente metafórica a la manera de la conducta de la rana, tenemos nuestras auténticas creencias y deseos y las acciones que elegimos ejecutar sobre la base de esos deseos y creencias. Cuando explicamos el hecho de que Mary corra escaleras arriba de repente citando su creencia de que se olvidó el bolso sobre la cama y su deseo de llevar el bolso consigo cuando salga, no estamos hablando metafóricamente; y si por casualidad hablamos de manera imprecisa (¿no creía ella en verdad simplemente que había dejado su bolso *en alguna superficie plana del dormitorio*?), esto siempre se puede corregir en principio, puesto que hay un hecho definido —dicen los realistas— acerca de qué contenido exacto tienen sus creencias y deseos.

Tenga en cuenta el ejemplo de la agenda de Quine:

> Hay un hombre de sombrero marrón a quien Ralph ha visto varias veces en circunstancias dudosas de las cuales no necesitamos ocuparnos aquí; basta decir que Ralph sospecha que es un espía. Hay también un hombre de cabello gris vagamente conocido por Ralph como pilar de la comunidad a quien Ralph no tiene conciencia de haber visto más que una vez en la playa. Ralph no lo sabe, pero los hombres son uno solo (1956, pág. 179).

Hay una enorme diferencia entre que Ralph crea la proposición de que el hombre del sombrero marrón es un espía y su creencia de la proposición de que el hombre a quien conoce como Ortcutt es un espía, ni siquiera cuando el hombre a quien conoce como Ortcutt resulta ser el hombre del sombrero marrón. Todos los matices de significado capaces de expresión de nuestro idioma pueden en principio distinguir los distintos deseos y creencias humanas. Como dice Davidson: "Sin el habla no podemos hacer las distinciones delicadas entre los pensamientos que son esenciales para las explicaciones que podemos a veces brindar con confianza. Nuestra manera de atribuir actitudes asegura que todo el poder expresivo del lenguaje se puede usar para hacer

esas diferencias". (1975, pág. 15-16.) Según este criterio, sigue siendo cierto, sin embargo, que en el estado común de las cosas, grandes familias de creencias viajan juntas por nuestras vidas mentales. (En un momento Mary cree que su bolso está sobre la cama, *y* cree que su bolso está sobre alguna superficie horizontal, y cree que el ítem que contiene su peine está sostenido por el mueble en el que duerme, etc., sigue una lista larga, tal vez infinitamente larga de otras creencias contemporáneas.)

De acuerdo con este criterio, las creencias y deseos son *actitudes proposicionales* y, por tanto, son tan numerosas y diferentes como las proposiciones que están disponibles como completadores de los objetos de las actitudes. Las ranas —o por lo menos los paramecios— no tienen, estrictamente hablando, actitudes proposicionales. Esto es lo que revelan los problemas de soltura y torpeza de ajuste cuando intentamos precisar sus creencias y deseos. Pero nosotros los seres humanos sí las tenemos y la "psicología de la actitud proposicional" es psicología adecuadamente así llamada.

El capítulo siguiente explora en detalle los apuros que acosan a cualquiera que tome este sendero realista. Aquí no hago más que señalar el momento cuando pienso que se toma la dirección equivocada: ocurre cuando se acepta el contraste implícito. Mi opinión es que la creencia y el deseo son como la creencia y el deseo propios de la rana ***hasta el fin***. Nosotros, los seres humanos, somos los sistemas intencionales más prodigiosos del planeta, y las enormes diferencias psicológicas entre nosotros y las ranas están mal descritas por el contraste propuesto entre la atribución de creencia literal y metafórica.

Esta polarización equivocada es una ilusión nacida del hecho de que no estamos sumergidos en información de la misma manera que la rana. No sólo saltamos, esquivamos, caminamos y comemos. Afirmamos, negamos, pedimos, ordenamos y prometemos. Y, además de nuestras actividades exteriores de comunicación pública, tenemos nuestras vidas contemplativas altamente verbales, en las cuales consideramos y formulamos hipótesis y distinguimos y ensayamos. Cuando no estamos hablando con otros, hablamos con nosotros mismos. Estas palabras en las que estamos sumergidos son las de nuestros idiomas naturales, tales como el inglés y el chino.

Si hay también o no un lenguaje del pensamiento, un medio simbólico más básico hecho realidad en nuestro sistema nervioso y lo bastante parecido a un lenguaje como para merecer ese nombre, es una pregunta concreta. Una de las fuentes poderosas de inspiración para la hipótesis del lenguaje del pensamiento es una ilusión que puede surgir de no poder distinguir entre las dos olas: la de información en la que tanto nosotros como las ranas estamos inmersos, y la de palabras de las que ninguna criatura, excepto la humana, tiene conciencia. Teniendo en cuenta la ubicuidad de las palabras, y nuestro incesante trabajo, juego y manoseo de las palabras, hay una provisión inagotable y en aumento permanente de artefactos humanos compuestos por palabras: no sólo las expresiones e inscripciones públicas, sino también las oraciones que nos pasen por la cabeza para que reflexionemos acerca de ellas, las aprobemos, desechemos, neguemos, memoricemos, admitamos. Estos productos de la actividad humana se confunden fácilmente con las cre-

encias (los deseos y otros estados mentales). (Véase "How to Change Your Mind" en *Brainstorms.*) Es decir, que existe una fuerte y pocas veces resistida tentación de suponer que, al identificar uno de estos estados, actos, productos verbalmente contaminados, hemos identificado un "estado mental interno que revela intencionalidad", subyacente y que estar en él explicaría lo que tomamos por tener creencias cuando nos entregamos a la psicología popular.

¿Qué cree Ralph acerca de Ortcutt? Si supusiéramos que Ralph fuera un fox terrier o un niño pequeño, estaría claro que sea lo que fuere lo que Ralph creyera acerca de Ortcutt, no será *acerca* de Ortcutt *qua* un hombre llamado Ortcutt, y que no será referido a que es un espía. El concepto de espía depende tanto de su papel en una sociedad verbal como el concepto de jueves o de apellido. Pero supongamos (como lo hace Quine) que Ralph es un adulto que utiliza el lenguaje, que vislumbra al furtivo Ortcutt y queda en cierto modo galvanizado y llevado a actuar por este hecho perceptivo. ¿Es esto semejante a la rana que salta hacia la izquierda? Por cierto que el contenido "total" de su estado perceptivo de "creencia" es tan resistente a la especificación precisa en términos de actitudes propositivas como su contrafigura en la rana, pero puede parecer que podemos extraer algunas proposiciones críticas para completar las atribuciones de creencia que nos interesan. Esto es solamente porque la acción de Ralph no es saltar hacia la izquierda (si esto fuera todo lo que se vio provocado a hacer, consideraríamos a todo el incidente como nada más que psicología de la rana) sino tender la mano hacia el teléfono, o apuntar una nota, o tal vez decir para sí mismo algo que conecta al hombre vislumbrado con el concepto de espionaje transportado por la palabra.

¿Y qué ocurre si Ralph no es impulsado en esa oportunidad a relacionarse de algún modo con cualquier producto verbal explícito semejante, pero a quien su experiencia expone no obstante a un estado de relativa cautela específica *vis-à-vis* Ortcutt? ¿Se puede catalogar *ese* estado por la vía de las atribuciones de actitud proposicional precisas? Supongamos que es un estado lo bastante definido como para controlar una voluminosa respuesta verbal de Ralph si se lo interrogara. "¿Qué piensa de ese hombre de sombrero marrón?", le sonsacaría expresiones explícitas de proposiciones abundantes en inglés, entre ellas la aseveración "el hombre es un espía". Una cosa que ocurre cuando se nos interroga acerca de lo que nosotros mismos creemos es que las oraciones de nuestros idiomas naturales surgen en nosotros como candidatas a ser aprobadas y posiblemente expresadas en público.

Se considera que este fenómeno es (equivalente a) la inspección introspectiva directa de las creencias de alguien. "¿Cómo puedo decir lo que pienso hasta que no veo lo que digo?", preguntó E. M. Forster. Hay mucho que decir acerca de esta maravillosa observación (véase *Brainstorms*, capítulo 16) pero aquí hay un punto equivalente: si bien es cierto, sin duda, que no hay en general ninguna manera mejor de expresar lo que alguien (incluido uno mismo) cree, que ver lo que dice, si uno considera los indicios que recibe de ese modo según el modelo de, digamos, la publicación de un poema o la liberación (por la Biblioteca del Vaticano) de un volumen retirado hasta ahora, uno podría muy bien estar cometiendo un error al suponer que un resfriado de cabeza está compuesto por un gran conjunto de estornudos internos, al-

gunos de los cuales se escapan. El proceso de examen de conciencia que individualiza tan claramente las expresiones de creencia no tiene por qué estar revelando ninguna individualización (presumiblemente de creencias) subyacente, psicológicamente importante, sino nada más que un artefacto de la exigencia ambiental de un tipo de acto determinado. (Churchland, 1981, pág. 85, ofrece una explicación especulativa de la declaración como una "*proyección* unidimensional, a través de las lentes compuestas de las áreas de Wernicke y Broca sobre la superficie de la idiosincrasia del lenguaje de quien habla una proyección unidimensional de un 'sólido' de cuatro o cinco dimensiones que es un elemento en su estado cinemático verdadero". Véase también Rosenberg, 1987.)

Nadie confunde recitar un credo con creer lo que el credo expresa y nadie confunde pronunciar una oración gramatical para uno mismo con creerla. ¿Entonces por qué las iglesias y las naciones le dan tanta importancia a hacer que la gente ejecute estos actos? Porque aunque tal expresión está conectada con la creencia sólo en forma indirecta, está poderosamente conectada con ella. Lo que la Iglesia y el Estado esperan lograr es inculcar la creencia que está "detrás" de la expresión. Quieren convertir la mera expresión oral en un juicio sincero. ¿Entonces hay también algún juicio que no sea sincero? Supongamos que uno "juzgue" en la cabeza. Se supone comúnmente que éste es un acto que manifiesta de manera muy directa una creencia, que tal vez origina un estado de creencia, que tal vez *es* una "creencia corriente". ¿Se puede formular un juicio sin creerlo? ¿Existe un fenómeno tal como un juicio hipócrita? (Un enigma de la psicología popular.) De cualquier modo podemos coincidir en que lo que explica las acciones de alguien no es el estado periférico de haberse relacionado con el producto público de un lenguaje sino el estado de creencia más profundo. No explicamos el hecho de que Lulú compre la *lasagna* con sólo citar el hecho de que escribió antes *"lasagna"* en su lista; ella tiene que tener las creencias y deseos pertinentes para apoyar su interpretación de lo que está inscrito en su lista de la compra. Y supongamos que ella memorizó la lista; todavía tiene que interpretar las palabras que le pasan por la cabeza como lista de la compra, y para eso necesita esas mismas creencias y deseos. ¿Cuáles son? Todavía más objetos de la lingüística, pero esta vez en el idioma mental, no en inglés, como dirían quienes creen en el idioma del pensamiento.

La hipótesis del lenguaje del pensamiento no niega el hecho evidente de que a veces tenemos en verdad oraciones y frases del *lenguaje natural* desfilando en la cabeza; afirma que detrás de la producción de oraciones en el lenguaje natural (en la cabeza o en público) hay todavía más oraciones en un idioma fundamental, no adquirido. Como veremos (en los dos próximos capítulos) hay buenas razones para querer hablar de un medio sistemático de representación en el cerebro, pero la idea de que los elementos de ese medio, los vehículos del significado deben reunir contenido en la clase de conjuntos típicos de las oraciones del lenguaje natural necesita un sustento independiente. Sin ese sustento, ¿qué razón hay para suponer que la creencia humana es tan diferente de la de la rana? En ambos casos se controla la conducta por medio de un complejo estado interno al que se puede ***aludir*** más o menos

eficazmente mediante las prácticas populares cotidianas de atribución de creencia y de deseo. Si en algún caso parece que las creencias se individualizan más fina y precisamente por sus contenidos, puede ser porque no les estamos prestando atención a las "creencias individuales" mismas (puede no existir tal esquema de individualización), sino a los productos de conductas lingüísticas controlados por esos complejos estados internos cuyos productos son *ipso facto* tan diferentes como lo permiten las discriminaciones en ese lenguaje.

Considérese uno de los ejemplos de Fodor de la doctrina del "Realismo Estándar": "saber que John cree que Mary lloró es saber que es muy probable que él piense que alguien lloró. Saber que Sam piensa que está lloviendo es saber que es muy probable que él piense que o está lloviendo o que John se fue y Mary lloró" (1985, pág. 87). El uso del verbo "pensar" a lo largo de todo este ejemplo es, a mi juicio, un síntoma de la confusión del Realista. No es improbable afirmar que si se puede hacer que John piense (en el sentido corriente de hablar con "uno mismo") que "Mary lloró" se puede hacer que sin ninguna enseñanza ulterior piense (asienta con la oración) "Alguien lloró". Como doctrina acerca de lo que es probable que la gente ***haría*** en el departamento del pensamiento ante distintas provocaciones, esto tiene cierta credibilidad. Como doctrina acerca de qué estados internos diferentes e individualizados un teórico equipado en forma adecuada encontraría al escudriñar el centro de creencia del sujeto, no tiene *ninguna credibilidad*.

Si interrogáramos a Ralph durante el tiempo suficiente podríamos obtener algunas declaraciones que sería mejor pasar por alto como indignas de confianza o indicadores engañosos del estado de Ralph. Tal vez éste, sin conocer a fondo los términos "insidioso" y "agente doble", hace algunas afirmaciones que nos informan mal acerca de la cuna de sus expectativas, de su capacidad deductiva, y así sucesivamente. Aun si Ralph dominara su vocabulario, sus tentativas para expresarse, como decimos, pueden ser indiferentes o poco juiciosas.

Según el punto de vista Realista, debe haber un hecho —por más difícil de determinar que sea— acerca de exactamente qué contenido (lo que significa: exactamente cuál proposición) se ha de encontrar en la creencia de Ralph. Los enigmas propios de los artefactos que surgen cuando uno trata de articular un conjunto de principios coherentes y psicológicamente plausibles para caracterizar este contenido han generado una industria casera de la teoría filosófica que ha chapuceado tanto que un grupo de universidades vecinas ha tenido durante años un "grupo de tareas de la actitud proposicional" muy presionada para no quedarse atrás con la literatura. Sin embargo, hace poco tiempo ha habido algunos cambios de parecer notables al aumentar las presiones. Como Loar (por aparecer) observa: "Ahora me parece algo extraordinario que hayamos podido pensar que los estados psicológicos están atrapados por un conjunto nítido de especificaciones de contenido".

Estoy sinceramente de acuerdo entonces con Stich y los Churchland en que el ingenio de esta reciente efusión nos recuerda el florecimiento tardío de los epiciclos ptolomeicos. Sin embargo, la orientadora visión realista de la "psicología de la actitud proposicional de los seres humanos" está todavía en severo contraste con lo que parece cómodo reconocer en el caso de la rana co-

mo sistema intencional: una caracterización de la rana desde la actitud intencional es siempre una idealización, y cualquier idealización se adaptará, sólo a medias, a los hechos brutales en el nivel físico de diseño. Más allá de los hechos acerca de exactamente dónde y cómo la mejor aproximación al nivel intencional resulta ser engañosa, simplemente no existen datos acerca de lo que la rana "realmente cree". La estrategia, aplicada a una rana, no necesita ni permite esa clase de precisiones. [1]

Algunos pueden no estar convencidos todavía de que la misma moraleja se aplica a nosotros mismos. Algunos pueden todavía tener la esperanza de salvar una teoría Realista de las actitudes proposicionales ***humanas***. Los dos capítulos siguientes y las reflexiones que los siguen deberían frustrar esa esperanza y reemplazarla con un sentido más Realista (si bien menos Realista) de lo que la psicología académica podría pensar de la psicología popular.

[1] Los límites de precisión en la atribución de estados intencionales a la rana han sido examinados recientemente en forma detallada, considerable y precisa por Israel (inédito). En los años transcurridos desde el trabajo clásico "What the Frog's eye Tells the Frog's Brain" por Lettvin, Maturana, McCulloch y Pitts (1959) los filósofos han representado con frecuencia a ranas (y sapos) en sus análisis del contenido mental. Es instructivo comparar las discusiones en Dennett (1969, págs. 48, 76-83), Stich (1981), Millikan (1986), y el capítulo 8 de este volumen con la explicación mucho más complicada de Israel, un tipo de "realismo naturalizado" presentado como alternativa del Realismo de Fodor y de Ewert (por aparecer), una presentación detallada de lo que se puede decir en forma corriente desde la actitud intencional acerca de la rapiña en el sapo.

5

Más allá de la creencia*

Supongamos que queremos hablar de las creencias. ¿Por qué queríamos hablar de ellas? No sólo "porque están ahí", puesto que está lejos de ser evidente que *están* ahí. Las creencias tienen una posición menos segura en una ontología científica crítica que, digamos, los electrones o los genes, y una presencia menos fuerte en el mundo cotidiano que, digamos, los dolores de muela o los cortes de pelo. Dar fundamentos para creer en las creencias no es un ejercicio justificado, pero tampoco es imposible. Una razón plausible y conocida para querer hablar de las creencias sería: porque queremos explicar y predecir la conducta humana (y animal). Es una razón tan buena como cualquier otra para querer hablar de las creencias, pero puede no ser suficientemente buena. Puede no ser suficientemente buena porque cuando uno habla de las creencias se implica en una maraña de problemas filosóficos de los cuales puede no haber escapatoria, excepto dejar de hablar de las creencias. En este ensayo trato de desenmarañar, o al menos exponer, algunos de estos problemas y sugerir formas en las que podríamos salvar algunas versiones o sustitutos teóricamente interesantes y útiles del concepto de creencia.

He aquí un resumen de la expedición que nos espera. Primero enfoco el problema principal: no bien está ampliamente aceptado que las creencias son *actitudes proposicionales* no hay ninguna interpretación constante y aceptada de ese término técnico. En la próxima sección, "Actitudes proposicionales", describo varias doctrinas incompatibles acerca de las proposiciones y por tanto acerca de las actitudes proposicionales que han sido desarrolladas como respuestas a las exigencias de Frege acerca de las proposiciones o "Pensamientos". Recientemente Putnam, y otros, han presentado ataques, con un tema común a todas las versiones de la doctrina estándar. En "Actitudes oracionales" discuto la retirada ante estos ataques que lleva (como lo hacen otras consideraciones teóricas) a postular un "lenguaje del pensamiento", pero muestro que hay serios problemas no resueltos en esta posición. En "Actitudes nocionales" esbozo un término medio alternativo, en efecto, entre las actitudes proposicionales y las oracionales, ni totalmente sintáctico ni totalmente semántico. Implica postular una ficción teórica: el *mundo nocional*. En la última sección "*De re* y *de dicto* desmantelados", las reflexiones previas produ-

* Publicado originalmente en *Thought and Object*, A. Woodfield, comp. (Oxford: Clarendon Press, 1982) y vuelto a imprimir con autorización.

cen diagnósticos alternativos de la panoplia de intuiciones conjuradas por la literatura *de re/de dicto*. Estas observaciones nos conceden la perspectiva de avanzar bien sin nada que se pudiera llamar en forma adecuada a la distinción entre las creencias de *re y de dicto*.

Recorreré un territorio muy conocido y virtualmente todo lo que voy a decir ya lo ha dicho, a menudo, mucha gente, pero considero que mi colección especial de temas conocidos y tal vez el orden e importancia que le doy, arrojará una nueva luz sobre los enigmas notablemente flexibles que han surgido en la literatura filosófica acerca de la creencia: enigmas acerca del contenido de la creencia, la naturaleza de los estados de creencia, la referencia o *acerquidad* en la creencia y la supuesta distinción entre las creencias *de re* y *de dicto* (o *relacionales y nocionales*). No propondré ni defenderé explícitamente ninguna "teoría" de la creencia. Todavía no he descubierto para qué serviría semejante teoría filosófica, lo que da lo mismo puesto que, de todos modos, sería totalmente incapaz de elaborar una. Este ensayo es más bien exploratorio y diagnóstico, un preludio, con suerte, de las teorías empíricas de los fenómenos que ahora comúnmente discutimos en términos de creencia.

Si entendemos que el proyecto ha de ser aclarar el concepto de creencia, todavía hay varias maneras distintas de concebirlo. Una manera es considerarlo como una parte pequeña pero importante de la semántica del lenguaje natural. Las oraciones con "...cree que..." y fórmulas parecidas a ellas son items que ocurren con frecuencia en el idioma natural inglés, y es así como uno quisiera reglamentar las presuposiciones e implicaciones de su uso, del mismo modo en que uno lo haría para otras expresiones del lenguaje natural, tales como "ayer" o "mucho" o "algunos". Muchos de los filósofos que contribuyen a la literatura que habla de los contextos de creencia, consideran que están haciendo exactamente eso —exactamente eso y nada más— pero en el transcurso de la exposición de sus propuestas soluciones a los enigmas conocidos de la teoría semántica de "cree" en inglés, aluden a doctrinas que los colocan de buen o mal grado en un proyecto diferente: defender (es decir, defender como *verdadera*) una teoría *psicológica* (o por lo menos un esbozo de teoría) de las creencias consideradas como estados psicológicos. Esta desviación de la semántica del lenguaje natural (o en cuanto a eso el análisis conceptual o la filosofía del lenguaje ordinario) a la metateoría de la psicología es natural, si bien no completamente inevitable, tradicional si bien no completamente ubicua y hasta es defendible, siempre que uno reconozca la desviación y asuma las cargas adicionales del argumento mientras aborda los problemas metateóricos de la psicología. Así es como Quine, Putnam, Sellars, Dummett, Fodor y muchos otros tienen observaciones que hacer acerca de cómo debe *proseguir* una teoría psicológica de la creencia. No tiene nada de malo sacar esas conclusiones del propio análisis del concepto de creencia, siempre que uno recuerde que si sus conclusiones son firmes, y ninguna teoría psicológica *andará* como uno llega a la conclusión de que *debe*, la conclusión ulterior correcta a sacar es: tanto peor para el concepto de creencia.

Si todavía queremos hablar de las creencias (mientras esperamos ese descubrimiento), tenemos que tener alguna forma de seleccionarlas o referirnos a ellas o distinguir unas de otras. Si las creencias son reales —es decir, es-

tados psicológicos reales de la gente— debe haber una cantidad indefinida de maneras en las que referirse a ellas, pero para ciertos fines algunas maneras serán más útiles que otras. Supongamos, por ejemplo, que fueran los libros, más que de las creencias, de lo que deseáramos hablar. Uno puede elegir un libro *por su título o su texto, o por su autor, o por su tema* o por la colocación física de una de sus copias ("el libro rojo está sobre el escritorio"). Lo que cuenta como *el mismo libro* depende en cierto modo de nuestros intereses del momento: a veces queremos decir *la misma edición* ("me refiero por supuesto al Hamlet de *First Folio*"); a veces nos referimos meramente al mismo texto aparte de algunos errores o correcciones); a veces meramente el mismo texto o una buena traducción (de otro modo, ¿cuántos de nosotros podrían afirmar haber leído algún libro de Tolstoi?). Esta última concepción de un libro es en algunos sentidos privilegiada: es lo que *habitualmente queremos decir* cuando hablamos sin salvedades especiales o sin indicaciones contextuales, acerca de los libros que hemos leído o escrito o queremos comprar. Usamos "una copia de..." para prefijar la noción de un libro en transacciones más concretas. Usted podría refutar a su oponente con *World and Object*, o si eso fallara darle un golpe en la cabeza con una *copia de... Being and Time*.

Hay un lugar parecido para la variación al hablar de creencias, pero la manera privilegiada de referirse a las creencias, lo que habitualmente queremos decir o creen que decimos en ausencia de salvedades especiales o indicaciones contextuales, es la proposición *creída*: por ejemplo, la creencia de que la nieve es blanca, que es la *misma creencia* cuando la creen fulano, zutano y mengano, y también cuando la creen los franceses monolingües, aunque los rasgos particulares de creencia en fulano, zutano, mengano, Alphonse y los demás, como las copias personales de un libro con las puntas dobladas de las páginas o las manchas de tinta podrían diferir de toda clase de maneras que no serían de interés para nosotros, dados nuestros propósitos normales al hablar de las creencias (Burge, 1979).

Por lo común, a las creencias se las considera *actitudes proposicionales*. El término es de Russell (1940). Hay tres grados de libertad en la fórmula de la actitud proposicional: persona, tipo de actitud, proposición: x cree que p, o y cree que p; x cree que p o *teme* que p o *espera* que p; x cree que p o que q, y así sucesivamente. De manera que podemos hablar de una persona que cree muchas proposiciones diferentes, o de una proposición en la que creen muchas personas diferentes o hasta de una proposición "adoptada" diversamente por gente diferente, o por la misma persona en ocasiones diferentes: yo solía dudar de que p, pero ahora estoy seguro de que p. Hay otras maneras de referirse a las creencias, tales como "la creencia que hizo ruborizar a Mary" o "la creencia más discutible de McCarthy", pero éstas son parasitarias; uno puede pasar a preguntar después de esa referencia: "¿y cuál es esa creencia?" en la esperanza de conseguir una referencia *identificatoria*: por ejemplo, "la creencia que tenía Mary de que Tom conocía su secreto. Algún día (cree cierta gente) podremos identificar las creencias de manera neurofisiológica ("la creencia que está en la corteza de Tom con el rasgo físico F"), pero por ahora, por lo menos, no tenemos cómo escoger una creencia del mismo modo en que podemos escoger un libro por la descripción física de una de sus copias o rasgos característicos.

La ortodoxia del punto de vista de que las creencias son actitudes proposicionales persiste a pesar de un sinnúmero de problemas. Siempre ha habido problemas relativamente "puros" *de los filósofos*, acerca del status metafísico o las condiciones de identidad para las proposiciones, por ejemplo; pero, con el nuevo interés por la ciencia cognitiva, hay que agregar ahora los problemas *de los psicólogos*, acerca de las condiciones para la ejemplificación individual de los estados de creencia, por ejemplo. La tentativa de aprovechar la discusión sobre la actitud proposicional como medio descriptivo de la teoría empírica en psicología les da un tono saludable a las presunciones ortodoxas y está inspirado un replanteamiento por parte de los filósofos. Muy a tiempo, se podría agregar, puesto que es claramente inquietante observar el entusiasmo con el que los no filósofos de la ciencia cognitiva están adoptando ahora formulaciones de la actitud proposicional para sus propios propósitos, en la inocente creencia de que cualquier concepto tan popular entre los filósofos debe ser sólido, convenido y bien probado. Ojalá fuera así.

Las actitudes proposicionales

Si pensamos que una buena manera de caracterizar el estado psicológico de una persona es caracterizar sus actitudes proposicionales, debemos suponer entonces que una exigencia crítica para obtener la descripción psicológica *correcta* será especificar las proposiciones *correctas* para esas actitudes. Esto a su vez nos exige decidir qué es una proposición y, todavía más importante, tener alguna opinión constante acerca de qué se toma como dos proposiciones diferentes y qué como una. Pero no hay consenso sobre estos asuntos del todo fundamentales. En realidad hay tres características generales totalmente distintas de las proposiciones en la literatura.

1) Las proposiciones son *entidades parecidas a las oraciones*, construidas de partes de acuerdo con una sintaxis. Como las oraciones, las proposiciones admiten un margen de duda acerca de la diferencia en el tipo del símbolo; los símbolos de las proposiciones del mismo tipo de proposición se han de encontrar en la mente (o el cerebro) de los creyentes en la misma creencia. Debe ser este punto de vista acerca de las proposiciones al que se acude en el debate entre los científicos cognitivos sobre *las formas de representación mental;* ¿es proposicional toda la representación mental o hay algo imaginista o análogo? Entre los filósofos, Harman (1973-1977) es sumamente explícito en la expresión de este punto de vista de las proposiciones.

2) Las proposiciones son *grupos de mundos posibles*. Dos oraciones expresan la misma proposición siempre que sean ciertas en, exactamente, el mismo grupo de mundos posibles. Según este criterio las proposiciones mismas no tienen ninguna propiedad sintáctica, y no se puede decir que tengan actitudes o símbolos en una mente o cerebro, o en una página. Stanaker (1976, 1984) defiende este criterio de las proposiciones. Véanse también Field (1978) y Lewis (1979) para conocer otras buenas discusiones de este criterio a menudo discutido.

3) Las proposiciones son como *colecciones o disposiciones de objetos y propiedades en el mundo*. La proposición de que Tom es alto consiste en

Tom mismo y el atributo o propiedad "altura". Russell sostenía ese criterio, y Donnellan (1974), Kaplan (1973, 1978, 1980) y Perry (1977, 1979) defendieron hace poco tiempo versiones especiales de él. Se oyen ecos del tema en muchos lugares diferentes —por ejemplo, en teorías de correspondencia de la verdad que afirman: lo que hace cierta una oración es su correspondencia con una realidad "del mundo" —donde una realidad resulta ser una proposición cierta.

Lo que, a mi modo de ver, une este disímil grupo de criterios acerca de que *son* las proposiciones es un conjunto de exigencias clásicas sobre lo que las proposiciones deben *hacer* en una teoría. Las tres exigencias se deben a Frege, cuya idea de un *Pensamiento* es la columna vertebral de la actual comprensión ortodoxa de las proposiciones (véase Perry, 1977). Y la diversidad de doctrinas se debe al hecho de que estas tres exigencias no pueden ser satisfechas simultáneamente, como veremos.

Según el punto de vista fregeano, una proposición (un pensamiento fregeano) debe tener tres características definitorias.

(a) Es una *portadora del valor de la verdad* (final, constante, original). Si p es verdadero y q es falso, p y q no son la misma proposición. (Véase Stich 1978a: "Si un par de estados puede ser de tipo idéntico... si bien difieren en el valor de la verdad, esos estados no son creencias tal como las imaginadas habitualmente.") (Véase también Fodor, 1980).

Esta es una condición exigida por la opinión común sobre las proposiciones como el medio fundamental de transferencia de la información. Si uno sabe *algo* y lo comunica a otros (en inglés, francés o a través de un gesto, o haciendo un dibujo) lo que éste adquiere es *la proposición* que sabe el primero. Sin embargo, se puede estar a favor de la condición (a) sin estar de acuerdo con este punto de vista de la comunicación o la transferencia de información. Evans (1980) es un ejemplo.

(b) Está compuesta por *intensiones*, entendiéndose por intensiones *a la* Carnap como determinantes de la extensión. Las distintas intensiones pueden determinar la misma extensión; la intensión de "tres al cuadrado" no es la intensión del "número de planetas" pero ambas determinan la misma extensión. Las diferentes extensiones, sin embargo, "no se pueden determinar por medio de una intensión".

Decir que las intensiones determinan las extensiones no es decir que las intenciones son *medios* o *métodos* para calcular las extensiones. Evans, en conferencias que dio en Oxford en 1979, llamó la atención sobre la tendencia a caer en esta comprensión de las intensiones y sugirió que contribuye de manera verosímil a puntos de vista como el de Dummett (1973, 1975) en el sentido de que lo que se sabe cuando se conocen los significados (o intensiones) es algo parecido a una ruta o método de verificación o procedimiento. Si esta posibilidad es totalmente espuria es una pregunta abierta e importante.

La condición (b) es de doble filo. Primero, puesto que las intensiones,

por las que está compuesta una proposición fijan sus extensiones en el mundo, *de qué trata una proposición* es una de sus características definitorias. Si *p* trata de *a* y *q* no, *p* y *q* no son las misma proposición. Segundo, puesto que la extensión no determina la intensión, el hecho de que *p* y *q* sean ambas acerca de *a*, y que ambas le atribuyan *F* a *a*, no alcanza para demostrar que *p* y *q* sean la misma proposición; *p* y *q* pueden tratar de *a* "de distintas maneras" —se pueden referir a *a* por la vía de distintas intensiones.

(Para Frege, las condiciones (a) y (b) se unificaron mediante su doctrina de que una oración enunciativa en su totalidad tenía una extensión: la Verdadera o la Falsa. En otras palabras, el todo intensivo, el Pensamiento, determina una extensión tal como lo hacen sus partes, pero mientras, sus partes determinan intensiones, o conjuntos de intensiones, el Pensamiento mismo tiene como intensión lo Verdadero o lo Falso.)

(c) Es "captable" por la mente.

Frege no nos dice nada de en qué consiste captar un Pensamiento, y a menudo se le ha criticado por esto. ¿Qué misteriosa transacción entre la mente (o el cerebro) y un objeto platónico —el Pensamiento— se supone que hay? (Véase, por ejemplo, Fodor, 1975, 1980; Field, 1978; Harman, 1977.) Esta pregunta propone una excursión pesada hacia la metafísica y la psicología especulativa, pero dicha excursión se puede postergar si observamos, como Churchland (1979) exhorta a hacerlo, que el catálogo de los predicados de la actitud proposicional tiene una analogía tentadora con el catálogo de los predicados de la medida física.

...cree que *p*	...tiene una longitud métrica de *n*
...desea que *p*	...tiene un volumen en m^3 de *n*
...supone que *p*	...tiene una velocidad en m/s de *n*
...está pensando que *p*	...tiene una temperatura en grados K de *n*

La sugerencia de Churchland es que las implicaciones metafísicas, si las hay, de los predicados de la actitud proposicional son las mismas de las de los predicados de las medidas físicas.

> La idea de creer que *p* es una cuestión de estar en cualquier relación adecuada con una entidad abstracta (la proposición que *p*) no tiene, en mi opinión, nada que la recomiende más que lo que tendría la sugerencia paralela de que pesar 5 kg es en el fondo sólo una cuestión de estar en alguna relación apropiada con una entidad abstracta (el número 5). Para los contextos de este último tipo, por lo menos, la interpretación relacional es altamente procristiana (inflexible). Contextos como
>
> *x* pesa 5 kg
> *x* se mueve a 5 m/s
> *x* irradia a 5 joules/s
> se catalogan en forma de mayor posibilidad con contextos como:

x pesa muy poco
x se mueve velozmente
x irradia copiosamente

En los tres últimos casos, lo que sigue al verbo principal tiene una función *adverbial*, transparente. Sugiere que la misma función adverbial se cumple en los primeros casos también. La única diferencia es que usar términos en singular para el número en posición adverbial proporciona una manera más exacta, sistemática y útil de modificar el verbo principal, especialmente cuando dicha posición está abierta a la cuantificación (1979, pág. 105).

Esta interpretación de las actitudes proposicionales no disuelve por sí misma los problemas metafísicos de las proposiciones, como veremos, pero, al unir su destino al destino de los números en física, desarma la sospecha de que hay un problema *especial* de los objetos abstractos en psicología. Más aun, nos permite distinguir dos puntos de vista que a menudo confluyen. Se piensa con frecuencia que *tomar en serio las proposiciones* debe implicar en psicología afirmar que las proposiciones *juegan un papel causal* de algún tipo en los sucesos psicológicos. De este modo, se nos lleva a pensar, como lo hace Harman (1977), acerca de la *función* de las proposiciones en el *pensamiento*. Para que las proposiciones tengan esa función, deben ser concretas —o tener signos concretos— y esto lleva de manera inevitable a una versión del criterio (1): las proposiciones son entidades parecidas a oraciones. No se podría suponer que conjuntos de palabras posibles o disposiciones de cosas y propiedades estuvieran ellas mismas "en la cabeza", y que sólo algo que está en la cabeza podría jugar un papel causal en psicología. Uno podría, sin embargo, tomarse en serio las proposiciones sin comprometerse con esta línea; uno podría tomárselas tan en serio como los físicos toman a los números. En la interpretación de Churchland, la *función* de una proposición no es sino la de *denotar* un término en singular que completa el modificador "adverbial" en un predicado de la actitud proposicional, un predicado que queremos usar para caracterizar el pensamiento, o la creencia, de otro estado psicológico de alguien. Este *no* es el criterio según el cual los predicados de la actitud proposicional no tienen ninguna estructura lógica; por cierto que mantiene la promesa de que las relaciones formales entre las proposiciones, como las relaciones formales entre los números, se pueden explotar en forma útil para formar los predicados de una ciencia. Es la opinión de que las proposiciones son objetos abstractos útiles para "medir" los estados psicológicos de los seres. Esto nos deja abierta la posibilidad de probar o descubrir más adelante que cuando un ser tiene una actitud proposicional determinada, algo en esa persona refleja la "forma" de la proposición: por ejemplo, es de algún modo isomórfica u homomórfica con la (canónicamente expresada) cláusula que *expresa* la proposición en la oración de la actitud proposicional. No hay necesidad de, y no habría que presuponer ninguna versión de esta dura afirmación, sin embargo, como parte de una comprensión inicial del significado de los predicados de la actitud proposicional.[1]

[1] Field, (en una postdata a Field 1978, en Block 1980, vol. 2) ve que este punto de vista lleva en forma inevitable a la afirmación fuerte de que:

No lograr hacer esta distinción, y persistir en esa actitud, ha creado un frustrante problema de comunicación en la literatura. En un caso típico, surge un debate acerca de la *forma* de las proposiciones en algún contexto especial de la actitud proposicional: por ejemplo, ¿son las proposiciones en cuestión condicionales universalmente cuantificadas, o disyuntivas infinitamente largas, o hay autorreferencia dentro de las proposiciones? Lo que no se aclara es si el debate es sobre la forma real de las estructuras cerebrales internas (en cuyo caso, por ejemplo, la pesadez de disyuntivas infinitamente largas plantea un problema real) o si es más bien un debate solamente acerca de la forma lógica correcta de los objetos abstractos, las proposiciones requeridas para completar los predicados que están en discusión (en cuyo caso las disyuntivas infinitas no tienen por qué plantear más problemas que en física). Tal vez algunos participantes en el debate están engañados por una presunción tácita de que el punto no puede ser real ni sustancial a menos que sea un punto *directamente* acerca de la forma física de las estructuras (estructuras "sintácticas") en el cerebro, pero otros participantes en el debate entienden que esto no es así, y por tanto persisten en defender su posición en el debate sin reconocer la posibilidad viva de que los dos lados no se estén escuchando mutuamente. Para evitar este conocido problema me atendré explícitamente a la mínima interpretación de Churchland como base neutral de operaciones desde la cual explorar las perspectivas de las interpretaciones más fuertes.

Con esta concepción metafísicamente restringida de las actitudes proposicionales en la mente, podemos entonces definir la noción evasiva de Frege acerca de la captación en forma muy directa: las proposiciones son captables si y sólo si los predicados de la actitud proposicional son predicados de la teoría psicológica proyectibles, predictibles, de buen comportamiento. (Uno podría, con el mismo espíritu decir que el éxito de la física, con su confianza

---

La teoría de la medición... explica por qué se pueden usar números verdaderos para "medir" masa) (mejor aun: para servir de balanza para masa). Lo hace de la siguiente manera. Primero se citan ciertas propiedades y relaciones entre los objetos masivos; propiedades y relaciones que se pueden especificar sin hacer referencia a los números. Luego se prueba un teorema de representación: ese teorema dice que si algún sistema de objetos tiene las propiedades y relaciones citadas, hay entonces una representación cartográfica de ese sistema dentro de los números verdaderos que "preserva la estructura". En consecuencia, asignarles números reales a los objetos es una manera conveniente de discutir las relaciones más intrínsecas que tienen esos objetos, pero esas relaciones intrínsecas no exigen por sí mismas la existencia de números reales.

¿Podemos resolver el problema de Brentano acerca de la intencionalidad de las actitudes proposicionales de manera análoga? Para hacerlo tendríamos que postular un *sistema de entidades* (la cursiva es mía) dentro del creyente que estaba relacionado por la vía de una representación cartográfica preservadora de la estructura con el sistema de proposiciones. La "estructura" que semejante representación cartográfica tendría que preservar sería el tipo de estructura importante para las proposiciones; viceversa, la estructura lógica, y creo que esto significa que el sistema de entidades dentro del creyente puede percibirse como un sistema de oraciones, un sistema interno de representación (pág. 114).

Para los objetos masivos postulamos o aislamos "propiedades y relaciones". ¿Por qué no propiedades y relaciones en lugar de "un sistema de entidades" en el caso de los sujetos psicológicos? Sea lo que fuere lo que podría ser "medido" por los predicados de la actitud proposicional se mide en forma indirecta, y el que Field opte por un lenguaje del pensamiento para "explicar" el éxito de la medición proposicional (cuyo alcance todavía está sin diagramar) es una suposición prematura, no una implicación de su punto de vista de los predicados.

en los números como factores formadores de predicados ¡demuestra que los números son captables por los objetos y procesos físicos!) La razón de ser de esta versión de la captación es que la exigencia de Frege de que las proposiciones sean algo que la mente puede captar es equivalente a la exigencia de que las proposiciones *tienen importancia* para una mente; es decir, para un estado psicológico de un ser. Se supone que lo que una persona hace es una función de su estado psicológico; las variaciones en el estado psicológico deberían predecir variaciones en la conducta. (Eso es lo que se supone que un estado psicológico es: un estado en el cual una variación es decisiva para la conducta.)[2] Ahora bien, si los estados psicológicos de las personas varían directamente con las caracterizaciones de su actitud proposicional de manera que, por ejemplo, cambiar las actitudes proposicionales de alguien es cambiar su estado psicológico y que compartir una actitud proposicional con otro es ser psicológicamente parecido al otro de algún modo, entonces las proposiciones figuran sistemáticamente en una interpretación lúcida de la psicología de la gente, y una manera clara de expresar esto sería decir que las personas (o los perros o los gatos, si los hechos se presentan así) captan las proposiciones que figuran en los predicados psicológicos que les corresponden. Ninguna otra forma más maravillosa de "aceptar" objetos abstractos o sus sustitutos concretos resulta implícita —*hasta ahora*— por sostener la condición (c). No hay duda de que Frege tenía una idea más ambiciosa, pero esta versión más débil de la captación es lo bastante exigente como para crear el conflicto entre la condición (c) y las condiciones (a) y (b).

Muchos autores han ofrecido recientemente argumentos para probar que las condiciones (a-c) no se pueden satisfacer en conjunto; lo que se puede captar no puede ser, al mismo tiempo, un determinador de extensión o un portador final del valor de la verdad: Putnam (1975a), Fodor (1980), Perry (1977, 1979), Kaplan (1980), Stich (1978a). (Entre las muchas discusiones relacionadas con esto, véase especialmente McDowell, 1977 y Burge, 1979.)

En primer término está el notorio experimento de pensamiento de Frege acerca del Planeta Gemelo de la Tierra. En pocas palabras (puesto que no nos detendremos ahora a explorar las innumerables objeciones que se le han hecho), el caso imaginado es el siguiente: hay un planeta, el Gemelo de la Tierra, que es casi un duplicado de la Tierra, hasta el punto de contener réplicas o *Doppelgängers* de toda la gente, lugares, cosas, acontecimientos, que hay en la Tierra. Hay una diferencia: los lagos, ríos, nubes, tuberías de agua, bañeras, tejidos vivos... no contienen H_2O sino XYZ, algo químicamente distinto pero indistinguible en sus macropropiedades normalmente observables, del agua, es decir H_2O. Los terráqueos del Planeta Gemelo *llaman* "agua" a este líquido, por supuesto, al ser átomo por átomo nuestras réplicas (¡Pase por alto la elevada proporción de moléculas de agua en no-

[2] *Tener una gran nariz colorada* es un estado que puede tener un lugar destacado en la psicología de alguien, pero no es por sí mismo un estado psicológico. *Creer que se tiene una gran nariz colorada* es uno de los muchos estados psicológicos que acompañaría comúnmente el hecho de tener una gran nariz colorada y sin el cual el estado de tener una gran nariz colorada tendería a ser psicológicamene inerte (como tener un gran hígado colorado). (Esto sería simplemente ostentación y no pretende explicar las diferencias intuitivas entre los estados psicológicos y los demás estados de un ser.)

sotros, por el placer de la discusión!). Ahora bien, puesto que mi *Doppelgänger* y yo somos réplicas *físicas* (¡por favor, en bien de la discusión!), seguramente somos también réplicas *psicológicas*; ejemplificamos las mismas teorías por encima del nivel al que H_2O y XYZ se pueden distinguir. Todos tenemos entonces los mismos estados psicológicos. Pero allí donde mis crencias son acerca del agua, las crencias de mi *Doppelgänger* (aunque exactamente de la misma "forma"), no son acerca del agua, sino acerca del XYZ. Creemos proposiciones diferentes. Por ejemplo, la creencia que yo expresaría con las palabras "agua es H_2O" es *acerca del agua* y *verdadera*; su contraparte en mi *Doppelgänger*, que, por supuesto, él explicaría con los mismos sonidos, no es sobre agua sino acerca de lo que él llama "agua", es decir XYZ, y es *falsa*. Somos gemelos psicológicos pero no gemelos de la actitud proposicional. Las actitudes proposicionales pueden variar con independencia del estado psicológico, de manera tal que las proposiciones [entendidas "clásicamente" como reuniendo las condiciones (a) y (b)], no son captables. Como lo expresa Putnam, algo se debe conceder: o el significado "no está en la cabeza" o el significado no determina extensión.

Stich (1978a) señala que es instructivo comparar este resultado con una opinión parecida, pero menos drástica que se expresa a menudo sobre el estado de *"conocimiento"*. Se observa con frecuencia que en tanto que "cree" es un verbo psicológico, "sabe" no es —o por lo menos no puramente— un verbo psicológico puesto que *x sabe que p* implica la verdad de *p*, algo que en general debe ser externo a la psicología de *x*. Por tanto, se dice, mientras *se cree que p* puede considerarse un estado psicòlógico (o mental) puro, *saber que p* es un estado "mestizo" o "híbrido", en parte psicológico, en parte algo más epistémico. En este caso, es el componente verbal del predicado de la actitud proposicional el que vuelve todo el predicado psicológicamente impuro y no proyectable. (Obsérvese que *es* no proyectable: experimentos simples que implicaran decepción o ilusión demostrarían inmediatamente que "*x* apretará el botón cuando *x* sepa que *p* es un pronóstico menos confiable que, digamos: "*x* apretará el botón cuando *x* esté seguro de que *p*.") Lo que el experimento de Putnam sobre el pensamiento pretende demostrar, sin embargo, es que aun cuando el verbo es un verbo psicológico aparentemente puro, el mero hecho de que el componente proposicional (*cualquier* componente proposicional) deba reunir las condiciones (a) y (b) vuelve todo el predicado psicológicamente impuro.

El experimento de Putnam sobre el pensamiento dista de ser indiscutible. Tal como está se apoya en doctrinas dudosas de tipo natural y concepción rígida, pero algunas variaciones sencillas sobre su tema básico pueden evitar por lo menos algunas de las objeciones más comunes. Por ejemplo, supongamos que el Planeta Tierra Gemelo es exactamente igual a la Tierra salvo porque yo llevo la billetera en el bolsillo de la chaqueta, y la de mi *Doppelgänger* no está en el bolsillo de su chaqueta. Yo creo (de verdad) que tengo la billetera en el bolsillo de la chaqueta. Mi *Doppelgänger* tiene la creencia opuesta. La suya es falsa. La mía verdadera. La suya no se trata de lo que se trata la mía: viceversa, *mi* billetera. diferentes proposiciones, diferentes actitudes proposicionales, la misma psicología.

En todo caso, Kaplan (1980) ha presentado un argumento acerca de un

caso parecido, con una conclusión parecida, que es quizá más preciso, pues no confía en estar de acuerdo con experimentos extravagantes del pensamiento sobre universos casi idénticos o sobre intuiciones de los tipos naturales que Putnam debe invocar para sustentar la afirmación de que XYZ no es sólo "Otra clase de agua".

Kaplan cita a Frege (1956):

> Si alguien quiere decir hoy lo mismo que expresó ayer usando la palabra 'hoy", debe reemplazar esta palabra con "ayer". Aunque el pensamiento es el mismo, su expresión verbal debe ser distinta, de modo que el sentido, que de otro modo sería afectado por los distintos momentos de expresión, se reajuste.

Pero al continuar, observa lo que Frege omitió:

> Si uno dice "Hoy es un día hermoso" el martes y "Ayer fue un día hermoso" el miércoles expresa el mismo pensamiento, de acuerdo con el trozo citado. Sin embargo, uno puede claramente perder la noción de los días y no darse cuenta de que está expresando la misma idea [el pensamiento fregeano, nuestra proposición]. Parece entonces que los pensamientos no son portadores apropiados de significación cognitiva.

Perry da todavía otro argumento que será discutido después y hay, además, otros argumentos en la literatura ya citada. [3]

No quiero apoyar ninguno de estos argumentos aquí y ahora, pero también quiero resistir el impulso —que aparentemente pocos pueden resistir— de cavar las trincheras aquí y ahora y pelear hasta la muerte en el terreno provisto sobre el Planeta Tierra Gemelo, los Tipos Naturales y lo que Frege en Realidad Quiso Decir. Propongo ceder un poco de terreno y ver dónde nos lleva.

Supongamos que estos argumentos son sólidos. ¿Cuál es su conclusión? Una afirmación a extraer de Kaplan (con la que Putnam y Perry estarían de acuerdo, me imagino) es ésta: Si hay algún factor indicativo en mi pensamiento o creencia, tal como "ahora" u "hoy", la proposición con la que estoy "vinculado" —la proposición que llena el blanco en el predicado correcto de la actitud proposicional que se refiere a mí— puede depender de manera crucial (pero imperceptible para mí) de sucesos tales como mover la manecilla de un reloj en el Observatorio de Greenwich. Pero es francamente increíble suponer que mi estado psicológico (mi estado de conducta vaticinado) pueda depender no solamente de mi constitución interna en el momento, sino por lo menos también de rasgos tan casualmente remotos como la disposición de las partes de algún cronómetro oficial. Eso no significa decir que mi futura conducta y mi futura psicología no pudieran ser *indirectamente* una función de mis verdaderas actitudes proposicionales de vez en cuando, por más desconocidas que fueran para mí. Por ejemplo, si apuesto a un caballo o rechazo una

[3] Uno de los más simples y convincentes se debe a Vendler (en una conversación): suponga que durante un período de más de diez años yo creo que Angola es una nación independiente. Intuitivamente, ésta es una *constante* del estado psicológico —algo acerca de mí que no cambia— y sin embargo esta sola creencia mía puede cambiar su valor de verdad durante la década. Si tomamos en cuenta mi estado como una creencia duradera, no puede ser una actitud *proposicional*.

acusación bajo juramento, los efectos *de largo alcance* de esta acción sobre mí se podrían predecir con más exactitud desde la proposición que yo *en realidad* expresé (y hasta creí; véase Burge, 1979) que desde la proposición que yo, por así decirlo, creí estar expresando, o creyendo.[4] Sin embargo este conocimiento sólo acrecienta el contraste entre la "asequibilidad" variable o poco digna de confianza de las proposiciones y la asequibilidad constitutiva o inherente de... *¿qué?* Si las proposiciones en el molde fregeano son vistas por estos argumentos como psicológicamente inertes (por lo menos en ciertas circunstancias determinadas), ¿cuál es la finalidad más accesible, comprensible y efectiva del papel proposicional?

Actitudes oracionales

¿Con qué reemplazaremos a las proposiciones? La respuesta más impulsiva (a juzgar por la cantidad de simpatizantes que tiene) es: algo parecido a *oraciones en la cabeza*. (Véanse, por ejemplo, Fodor, 1975, 1980; Field, 1978; Kaplan, 1980; Schiffer, 1978; Harman, 1977; pero si busca reflexiones eminentes, véase Quine, 1969.) Al final encontraremos que esta respuesta no es satisfactoria, pero entender su llamamiento es, me parece, un preliminar esencial del trabajo de encontrar una retirada mejor de las proposiciones. Hay muchos caminos para llegar a las *actitudes oracionales*.

He aquí el más sencillo: cuando uno capta *figuradamente* una proposición, que es un objeto abstracto, debe captar *literalmente* algo concreto, pero de algún modo parecido a una proposición. ¿Qué podría ser esto excepto una oración en la mente o cerebro, una oración en el lenguaje mentalista? (Para aquellos que ya tienen el criterio de que las proposiciones son cosas parecidas a oraciones, ésta es, por cierto, una retirada corta; consiste en renunciar a las condiciones (a) y (b) para las proposiciones; pero entonces, si las proposiciones han de ser algo más que oraciones no interpretadas, ¿qué más son? Hay que poner algo en lugar de (a) y (b).

He aquí otro camino para llegar a las actitudes oracionales. ¿Cuáles son los verdaderos *constituyentes* de los estados de creencia sobre perros y gatos? No gatos y perros reales y vivientes, por supuesto... sino símbolos de, o representaciones de perros y gatos. La creencia de que el gato está en la alfombrilla consiste de algún modo en una representación estructurada compuesta de símbolos para el gato y la alfombrilla y la relación *en* —una especie de oración— o quizás una especie de cuadro[5] ante el cual el observador dice "¡Sí!". No se trata de que la creencia no tenga eventualmente que unir el creyente con el mundo, con perros y gatos reales y vivos, pero el problema de esa relación de unión puede ser aislado y postergado. Se lo convierte en el problema aparentemente más tratable y conocido de la *referencia* de los tér-

[4] Se podría resumir el caso que Putnam, Kaplan y Perry presentan de este modo: las proposiciones son *inasibles* porque pueden *eludirnos*; la presencia o ausencia de una proposición determinada "en nuestro poder" puede ser psicológicamente improcedente.

[5] Los temas que se consideran en las controversias palabras-mentales-versus-imágenes mentales son mayormente ortogonales a los puntos que se discuten aquí, que se refieren a problemas que deben ser resueltos antes de que *ya* a las imágenes mentales o a las oraciones mentales se les pueda extender un certificado de salud limpio como entidades teóricas.

minos en las oraciones, en este caso oraciones mentales internas. Uno imagi na que los grandes caballos de batalla de la educación lógica-Frege, Carnap Tarski— pueden ser ensillados de inmediato para esta tarea. (Véase, Field 1978, para la defensa más explícita de este camino.)

He aquí el tercer camino. Necesitamos una explicación física, causal del fenómeno de la opacidad; el hecho de pensar que sería bonito casarse con Yocasta es un estado con distintas consecuencias psicológicas, con diferentes efectos en el mundo; a partir del estado de creer que sería lindo casarse con la madre de Edipo, a pesar de la hoy bien conocida identidad. Una sugerencia tentadora es que estos dos estados diferentes, en sus realizaciones físicas en un creyente, tienen en efecto una sintaxis, y la sintaxis de un estado se parece a, y es diferente de la sintaxis del otro exactamente en la manera en que las dos oraciones de atribución se parecen y difieren; y que los efectos diferentes de los dos estados se pueden rastrear eventualmente hasta estas diferencias en la estructura física. Esto se puede considerar como una explicación de la opacidad del discurso *indirecto* incluyéndolo en la *super* opacidad de algo del discurso *directo*: la cita estricta, en efecto, de distintas oraciones en el lenguaje mentalista (Fodor, 1980).

He aquí, finalmente, el camino hacia las actitudes oracionales más pertinente a los problemas que hemos descubierto con las actitudes proposicionales. Aparentemente, lo que hace que el ejemplo de "hoy" y "ayer" de Frege-Kaplan funcione es lo que podría llamarse la impermeabilidad de las proposiciones a los índices. Esta impermeabilidad se explicita en el sustituto de Quine para las proposiciones, las *oraciones eternas*, que están equipadas cada vez que se las necesite con variables limitadas de tiempo, espacio y persona para extirpar los efectos variables o de perspectiva de los índices (Quine, 1960). Una alternativa psicológicamente lúcida para las proposiciones resistiría precisamente este movimiento y construiría de algún modo elementos indicativos donde fueran necesarios. Ya hay un modelo obvio listo para manejar: las oraciones, ordinarias, externas concretas, oraciones expresadas de los lenguajes naturales, habladas o escritas. Las oraciones deben ser entendidas, ante todo, como objetos sintácticamente individualizados, como sartas de símbolos de "formas" determinadas, y una observación estándar de las oraciones tan individualizada que los signos de un tipo determinado de oraciones pueden "expresar" diferentes proposiciones según el "contexto". Los signos del tipo de oración "Estoy cansado" expresan diferentes proposiciones en bocas distintas en momentos distintos. A veces "Estoy cansado" expresa una proposición real acerca de Jones, y a veces una proposición falsa acerca de Smith. tal vez haya, como alega Quine, algunos tipos de oraciones, las oraciones eternas, cuyos signos expresan todos, en efecto, la misma proposición. (Quine —que no es amigo de las proposiciones— debe ser, en efecto, más cauto al formular esta afirmación.) Pero es precisamente el poder de las otras oraciones, las no eternas, de ser de contexto seguro, lo que se necesita, intuitivamente, para el papel de individualizar los estados y sucesos psicológicamente notables. La indicatividad de las oraciones parece ser la contrafigura lingüística de esa relatividad a un punto de vista subjetivo que es el sello distintivo de los estados mentales (Castañeda, 1966, 1967, 1968; Perry, 1977, 1979; Kaplan, 1980; Lewis, 1979).

Si *lo que significa una oración* fuera tomado por ser *la proposición que expresa*, entonces signos diferentes de un tipo de oración indicativa significarían cosas distintas, y sin embargo, todavía parece haber lugar y necesidad para un sentido del "significado" según el cual podemos decir que todos los signos de un tipo de oración *significan la misma cosa*. Un *tipo* de oración, hasta un tipo indicativo tal como "Estoy cansado", significa algo —una sola "cosa"— y de ahí que en ese sentido, ocurra lo mismo con todos sus signos. La misma cosa no es una proposición, por supuesto. Kaplan sugiere que la llamemos el "*carácter*" de la oración. "El carácter de una expresión está dado por convenciones lingüísticas y, a su vez, determina el contenido de la expresión en todo contexto." La sugerencia de Kaplan se despliega en una figura simétrica de dos etapas de la interpretación de la oración:

> Así como fue conveniente representar los contenidos por funciones desde las circunstancias posibles hasta las extensiones (las intensiones de Carnap), así es conveniente representar a los caracteres por medio de funciones desde los contextos posibles de la expresión hasta los contenidos... Esto nos da el siguiente cuadro:
>
> Carácter: Contextos → Contenido
>
> Contenido: Circunstancias → Extensiones
>
> o, en un lenguaje más conocido,
>
> Significado + contexto → Intensión
>
> Intensión + mundo posible → Extensión

Aunque Kaplan habla de oraciones públicas, externas —no de oraciones en la cabeza, en el lenguaje mentalista escrito— la pertinencia de esta clase de significado lingüístico a la psicología es aparente de inmediato. Kaplan comenta: "Puesto que el carácter es lo que está establecido por las convenciones lingüísticas, es natural pensar acerca de ella como *significado* en el sentido de lo que es conocido por el usuario competente del lenguaje. "Ningún dominio de mi lengua nativa asegurará que pueda decir *qué proposición* he expresado cuando articulo una oración, pero mi competencia como hablante nativo sí me da acceso, aparentemente, al *carácter* de lo que he dicho. ¿No podríamos generalizar el punto llevándolo al lenguaje postulado del pensamiento y tratar el carácter de las oraciones en lenguaje mentalista como lo que es captado directamente cuando uno "acaricia" ("articula" mentalmente) una oración en el lenguaje mentalista? ¿No podrían ser los *objetos* de la creencia los caracteres de las oraciones en lenguaje mentalista en lugar de las proposiciones que expresan?

Comentando a Kaplan, Perry (1977, 1979) desarrolla este tema. Donde Kaplan habla de *expresiones* del mismo *carácter*, Perry desplaza el punto dentro de la mente y habla de personas que *abrigan* los mismos *sentidos*, y allí donde Kaplan habla de *expresiones* que tienen el mismo *contenido*, Perry habla de personas que *piensan* el mismo *pensamiento*. (Los términos de Perry deliberadamente imitan a Frege, por supuesto.) Este expresa con precisión el llamamiento de este movimiento teórico en un pasaje también citado por Kaplan:

> Usamos los *sentidos* (los caracteres de Kaplan, en esencia) para individualizar los estados psicológicos al explicar y predecir la acción. Es el sentido tomado en consideración y no el pensamiento percibido, lo que está ligado a la acción humana. Cuando usted y yo consideremos el sentido de "Está a punto de atacarme un oso" nos comportamos de manera parecida. Los dos nos hacemos una bola y tratamos de estar lo más quietos posible. Distintos pensamientos comprendidos, el mismo sentido tomado en consideración, la misma conducta. Cuando tanto usted como yo comprendemos la idea de que estoy a punto de ser atacado por un oso, nos portamos de diferente manera. Yo me hago una bola, usted corre en busca de ayuda. El mismo pensamiento percibido, un sentido diferente tomado en consideración, una conducta diferente. Asimismo, cuando usted cree que la reunión comienza al mediodía de un día dado, por considerar el día anterior el sentido de "la reunión comienza mañana al mediodía", no hace nada. Al percibir la misma idea al día siguiente, al considerar el sentido de "la reunión empieza ahora" usted salta de la silla y corre por el pasillo (1977, pág. 494).

El concepto, entonces, es que postulamos un lenguaje del pensamiento, que tal vez sea completamente distinto a cualquier lenguaje natural que un creyente pueda saber, y adaptamos la explicación en dos etapas del significado (carácter + contenido) dada por Kaplan y que estuvo diseñada inicialmente para la interpretación semántica de las expresiones del lenguaje natural, como una explicación en dos etapas de la interpretación semántica de los estados psicológicos. La primera etapa, los *sentidos* de Perry (modelados sobre los *caracteres*) nos daría predicados psicológicamente puros con objetos captables; la segunda etapa, los *pensamientos* de Perry (modelados sobre los *contenidos* de Kaplan), serían psicológicamente impuros pero completarían el trabajo de la interpretación semántica llevándonos hasta el fin (por la vía de las intensiones carnapianas, en efecto) a las extensiones: las cosas del mundo acerca de las cuales son las creencias.

He aquí la propuesta desde otra perspectiva. Supongamos que hubiéramos comenzado con la pregunta: ¿Qué es lo que determina lo que un ser (cualquier entidad con estados psicológicos) cree? Es decir, ¿qué características de la entidad, considerada completamente sola y aislada de su inserción en el mundo, fijan las proposiciones de sus actitudes proposicionales? La asombrosa respuesta de Putnam a esta pregunta es: ¡nada! Todo lo cierto acerca de esa entidad considerada por sí misma es insuficiente para determinar su creencia (sus actitudes proposicionales). Los hechos acerca de la inserción ambiental / causal / histórica de la entidad —el "contexto de expresión" en efecto— se deben agregar antes de que tengamos suficiente para fijar proposiciones.

¿Por qué es asombrosa la respuesta? Debería ser asombrosa para cualquiera que tuviera un recuerdo cariñoso de las *Meditaciones* de Descartes, puesto que en las *Meditaciones* parecía firmemente seguro que el *único* tema que fue fijado y determinado exclusivamente dentro de los límites de la mente de Descartes fue exactamente los pensamientos y creencias que tenía en ese momento (qué proposiciones abrigaba). Descartes podría deplorar su incapacidad para decir cuáles de sus creencias o pensamientos eran verdaderos, cuáles de sus percepciones verídicas, pero cuáles pensamientos o creencias **eran**, la identidad de sus propios candidatos personales para la verdad y la

falsedad, parecía estar completamente determinado por la naturaleza interna de su propia mente, y además clara y netamente captable por él. Pero si Putnam, Kaplan y Perry están en lo cierto, a Descartes le iba peor de lo que pensaba; ni siquiera podía estar seguro de qué proposiciones consideraba.

Hay por lo menos cuatro maneras de resolver este conflicto. Se podría tomar partido por Descartes y buscar un rechazo convincente en la línea de pensamiento de Putnam. Se podrían aceptar las conclusiones putnamianas y rechazar a Descartes. Se podría observar que el punto de vista de Putnam no es un ataque directo a Descartes, puesto que presupone la condición física de la mente, lo que por cierto Descartes desaprobaría; se puede decir que Descartes podría conceder que todo lo *físico* acerca de mí y mi *Doppelgänger* indetermina nuestras actitudes proposicionales, pero que no obstante, ellas están determinadas "interiormente" por características de nuestras mentes no físicas, que deben ser de naturaleza lo bastante distinta como para fijar nuestras diferentes actitudes proposicionales. O se podría intentar un arreglo más conciliador, aceptando el punto de vista de Putnam contra las proposiciones entendidas como objetos que reúnen condiciones (a-c) y afirmando que lo que Descartes podía captar desde una posición de privilegio eran más bien los *verdaderos* objetos psicológicos de sus actitudes, no proposiciones sino los sentidos de Perry.

Al seguir la última corriente cambiamos nuestra pregunta inicial: ¿cuál es, entonces, la *contribución orgánica* a la fijación de las actitudes proposicionales? ¿Cómo caracterizaremos lo que obtengamos al sustraer los hechos del contexto o la inserción del todo determinante? Este residuo, como quiera que debiéramos caracterizarlo, es el campo de acción mismo de la psicología, la psicología "pura", o según la frase de Putnam, "psicología en el sentido estrecho". Focalizar la contribución orgánica de forma aislada es lo que Putnam llama *solipsismo metodológico*. Cuando Fodor adopta el término y recomienda el solipsismo metodológico como una estrategia de investigación en psicología cognitiva (1980), recomienda precisamente este movimiento.

¿Pero cómo se procede con esta estrategia? ¿Cómo caracterizaremos la contribución orgánica? Debería ser análoga a la noción de carácter de Kaplan, de manera que empezamos, tal como lo hizo Perry por psicologizar el esquema de Kaplan. Cuando lo intentamos notamos que el esquema de Kaplan es incompleto, pero que puede extenderse directamente en consonancia con sus comentarios de sustentación. Recuérdese que Kaplan sostenía que las "convenciones lingüísticas" determinan el carácter de cualquier tipo determinado de expresión. De manera que el proceso de interpretación en dos etapas de Kaplan va precedido, en efecto, por una etapa anterior (0), gobernados por las convenciones lingüísticas:

(0) Características sintácticas + convenciones linguísticas → Carácter

(1) Carácter + Contexto → Contenido

(2) Contenido + Circunstancias → Extensión

Cuando psicologicemos el esquema ampliado, ¿qué pondremos en lugar del primer espacio? ¿Cuál será nuestro término análogo de las características

sintácticas de las expresiones que, dadas las convenciones lingüísticas, fijan el carácter? Es aquí donde hace su entrada nuestro compromiso con un lenguaje del pensamiento. Las "expresiones" en el lenguaje del pensamiento se necesitan como la "materia prima" para la interpretación psicológico-semántica de los estados psicológicos.

Por supuesto, esto es exactamente lo que esperábamos y al principio todo parece andar muy bien. Considérese el experimento de Putnam sobre el pensamiento. Cuando él introduce un *Doppelgänger* o duplicado físico confía tácitamente en nuestro consentimiento de la afirmación de que, puesto que dos réplicas exactas —mi *Doppelgänger* y yo— tenemos exactamente la misma estructura en todos los niveles de análisis desde el microscopio para arriba; cualesquiera que sean los sistemas sintácticamente definidos que cualquiera de nosotros pueda encarnar, el otro también lo encarna. Si *Yo* pienso en escritura cerebral o en lenguaje mentalista, pensamientos con exactamente las mismas "formas" ocurren también en mi *Doppelgänger* y el corolario tácito ulterior es que mis pensamientos y los de mi *Doppelgänger* también serán de tipo de idéntico *carácter*, en virtud de su identidad de tipo sintáctico. Mis pensamientos, sin embargo, son acerca de mí, mientras que los de él son acerca de él —aunque los nombres que se da a sí mismo son sintácticamente los mismos que mis nombres para mí mismo: ambos nos llamamos "Yo" o "Dennett". El punto de vista de Putnam ilustra, entonces, aparentemente de manera muy precisa el esquema de Kaplan en acción. Mi *Doppelgänger* y yo tenemos pensamientos con el mismo carácter (sentido, para Perry), y todo lo que se necesita es un contexto —la Tierra o la Tierra Gemela— para explicar la diferencia en el contenido (la proposición según el pensamiento de Perry y Frege) expresado, y de allí la diferencia en extensión dada la circunstancia.

¿Qué presunciones, entonces, permiten el corolario tácito de que la identidad de tipo sintáctico es suficiente para la identidad del tipo de carácter? ¿Por qué no es posible que aunque un pensamiento en la *forma* de "estoy cansado" ocurre tanto en mí como en mi *Doppelgänger;* en *él* el pensamiento con esa forma significa *la nieve es blanca* y por tanto difiere no sólo en la proposición expresada sino también en carácter? Admito que en cualquier punto de vista sólido del idioma mentalista (si es que existe alguno), la identidad del tipo de carácter debería ser la consecuencia de la semejanza física, pero, ¿por qué? Se debe deber a una diferencia entre las personas y, digamos, los libros, puesto que una réplica átomo por átomo de *la autobiografía de Malcolm X* en cualquier otro planeta (o en cualquier otro lugar donde hablaran schmenglish) podría no ser la autobiografía de nadie; sería una monografía sobre lógica epistémica o una historia de las guerras. [6] La razón

[6] John McCarthy afirma que esto es demasiado fuerte: los patrones de repetición y co-ocurrencia puramente formales que se encuentran en listas de caracteres largas como un libro, ejercen una represión muy fuerte —a la que podríamos llamar la represión del criptógrafo— sobre cualquiera que trate de idear interpretaciones de un texto que no sean diferentes de manera trivial. Muchos "trucos baratos" producirán distintas interpretaciones de poco interés. Por ejemplo, declarar que la primera persona del singular en inglés es una variedad de la tercera persona del singular en schmenglish y (con algunos aspavientos) convertir una autobiografía en una biografía. O declarar que el schmenglish tiene palabras muy largas —en realidad tan largas co-

por la cual podemos pasar por alto esta alternativa en el caso del lenguaje mentalista debe tener que ver con una relación más íntima entre forma y función en el caso de cualquier cosa que pudiera pasar por lenguaje mentalista en contraste con un lenguaje natural. De manera que el papel jugado por las convenciones lingüísticas en la etapa (0) tendrá que ser jugado en la versión psicológica del esquema por algo que no es *convencional* para nada en ningún sentido ordinario.

En realidad, cuando nos volvemos a la tentativa de llenar los detalles de la etapa (0) para el esquema psicológico descubrimos una multitud de perplejidades. ¿De dónde vienen los rasgos sintácticos del lenguaje mentalista? La psicología no es hermenéutica literaria; el "texto" no está *dado* ¿qué formas de cosas en la cabeza cuentan? Parece que Kaplan ha pasado por alto otra etapa primordial en su esquema.

(−1) rasgos físicos + consideraciones de diseño (en otras palabras *menos* extrañezas funcionales)

Las diferencias en el tipo de letra, color y tamaño de los signos escritos y en el volumen, tono y timbre de los signos hablados no cuentan como diferencias sintácticas, excepto cuando se puede demostrar *que funcionan* como diferencias sintácticas marcando "valencias" combinatorias, posibilidades de variaciones del sentido, etcétera. Una caracterización sintética es una abstracción considerable de los rasgos físicos de los signos; los signos del código Morse produciéndose a tiempo pueden compartir su sintaxis con los signos de oraciones impresas en inglés.

Por analogía podemos esperar entonces que los signos de la escritura cerebral difieran en muchos rasgos físicos y que sin embargo cuenten como compartiendo una sintaxis. Nuestro modelo nos concede campo suficiente para declarar físicamente que "sistemas de representación" muy diferentes son meras "variantes notacionales". De cualquier modo ésta es una idea conocida que es sólo un caso especial de la libertad de realización física defendida por teorías funcionalistas de la mente (por ejemplo, Putnam, 1960, 1975b; Fodor, 1975). Seguramente el creyente en la psicología de la actitud oracional se sentirá muy agradecido por este campo de acción, puesto que la posición que atrae al final de este camino es asombrosamente fuerte. Llamémosla *oracionismo:*

(0) x cree lo que y cree si y sólo si (L) (s) (L es un lenguaje del pensamiento y s es una oración de L y hay un signo de s en x y un signo de s en y)

(Debemos comprender que estos signos tienen que estar en los lugares funcionalmente pertinentes y semejantes por supuesto. Uno no puede llegar a creer lo que Jones cree escribiendo sus creencias (en L) en tiras de papel y tragándoselas.)

---

mo un capítulo en inglés— y transformar cualquier libro de diez capítulos en la oración de diez palabras de su elección. La perspectiva de diferentes interpretaciones interesantes de un texto es difícil de evaluar pero digna de explorar, puesto que proporciona una condición límite para la "traducción literal" (y, por tanto, una "interpretación literal". Véanse "Los experimentos del pensamiento" de Lewis, 1974).

Desde este punto de vista debemos *compartir* un lenguaje del pensamiento para creer *la misma cosa* —aunque por "la misma cosa" ya no queremos decir la misma proposición. La idea es que, desde el punto de vista de la psicología, una asignación de papeles diferentes es más apropiada, una asignación según la cual las creencias resultan ser las mismas cuando tienen el mismo sentido (según Perry) o cuando consisten en relaciones con oraciones internas del mismo carácter. Sin embargo, al eludir las proposiciones, hemos renunciado a uno de sus rasgos útiles; la neutralidad lingüística. La exigencia de que interpretemos todo como creyentes (en la nueva asignación de papeles, psicológicamente realista) como pensando en el mismo lenguaje es pesada, aparentemente, a menos que, por supuesto, se pueda encontrar algún modo de defender esta implicación o de convertirla en una trivialidad.[7]

¿Cómo podríamos defender el oracionismo? Declarando que es una pregunta empírica, muy interesante e importante, además, acerca de si la gente piensa realmente en el mismo o en diferentes lenguajes del pensamiento. Tal vez los perros piensen en *perrés* y la gente en *gentés*. Tal vez descubramos la "escritura del pensamiento que la gente tiene en común independientemente de su nacionalidad y otras diferencias" (Zeman, 1963). Semejante descubrimiento sería sin duda un tesoro teórico. O quizá resultaran ser diferentes idiomas mentales poco triviales (no meras variantes notacionales mutuas) de manera tal que la implicación con la que tenemos que convivir es que si su cerebro habla latín mental mientras que el mío habla griego mental, es imposible que *compartamos los mismos estados psicológicos*. Cualquier comparación "perdería algo con la traducción". Esto no nos impediría, necesariamente, compartir actitudes *proposicionales* (ya no consideradas más estados psicológicos "puros"); su manera de creer que las ballenas son mamíferos sería sólo diferente —en modos psicológicamente no triviales— de la mía. Sería una calamidad teórica descubrir que cada persona piensa en una lengua mentalista diferente, completamente idiosincrática, puesto que entonces sería difícil obtener una generalización psicológica; pero si resultara haber un número menor de distintas lenguas del pensamiento (con el agregado de unos pocos dialectos), esto resultaría ser tan teóricamente fructífero como el descubrimiento monolingüe, puesto que podríamos explicar diferencias importantes en los estilos cognitivos mediante una hipótesis multilingüe. (Por ejemplo, a la gente que piensa en latín mental le va mejor en ciertas clases de problemas del razonamiento, mientras que la persona que piensa en griego mental son superiores como descubridores de analogías. Recuérdese todos los viejos dichos acerca de que el inglés es el idioma del comercio, el francés el idioma de la diplomacia y el italiano el idioma del amor.)

[7] Field (1978) nota este problema, pero, sorprendentemente, lo descarta: "La noción de identidad tipo entre los signos de un organismo y los de otro no es necesaria para la teoría psicológica y puede ser considerada como una noción sin sentido" (pág. 58, nota 34). Sus razones para sostener este punto de vista notable no son menos notables, demasiado tortuosas como para hacerles justicia explícitamente aquí. Hay muchos puntos de acuerdo y desacuerdo importantes entre el trabajo de Field y éste, más allá de aquellos que voy a discutir, pero discutirlos a todos duplicaría la extensión de este trabajo. Discuto el compromiso de Fodor (1975) con el oracionismo en "A Cure for the Common Code" en *Brainstorms*.

En cualquier caso, hay una estrategia del desarrollo teórico disponible que tenderá a reemplazar las hipótesis o interpretaciones multilingües con hipótesis o interpretaciones monolingües. Supongamos que hayamos fijado tentativamente un nivel de descripción funcional de dos individuos, según la cual no son colingües; sus estados psicológicos no comparten una sintaxis. Podemos tratar de encontrar un nivel algo más alto de abstracción en el cual podamos redescribir sus estados psicológicos de manera de que lo que había sido tratado hasta ahora como diferencias sintácticas pueda ser descartado ahora como diferencias físicas por *debajo de la sintaxis*. Al nivel funcional más alto descubriremos que *la misma función* está apoyada por lo que habríamos estado considerando con items sintácticamente diferentes, y esto nos dará derecho a declarar la taxonomía sintáctica anterior como demasiado fina (para distinguir las que son variantes meramente nocionales de sistemas rivales de signos). Valernos de esta estrategia tenderá a embromar las líneas entre la sintaxis y la semántica, puesto que dependerá de la capacidad de lo que consideramos como rasgo sintáctico a cierto nivel de análisis, figurar en diferencias semánticamente pertinentes al nivel siguiente más alto del análisis funcional. Llevado a un extremo pickwickiano podríamos encontrarnos a un nivel *muy* abstracto del análisis funcional defendiendo una versión de monolingüismo del lenguaje mentalista análogo a la siguiente afirmación acerca del lenguaje natural: el francés y el inglés son sólo variantes nocionales uno del otro o de algún *ur*-idioma; *bouche y mouth* (boca [T.]) son diferentes signos del mismo tipo (véase Sellars, 1974). Normalmente, y por buenas razones consideramos que estas dos palabras comparten sólo propiedades *semánticas*. Principios parecidos presumiblemente restringirían nuestra teorización acerca del lenguaje mentalista y sus dialectos. Quienquiera que crea que *tiene* que haber una sola lengua mentalista para los seres humanos debe estar pasando por alto la existencia de dichos principios y quedándose prendado de la versión pickwickiana del monolingüismo.

Aun si aceptamos el multilingüismo y lo encontramos teóricamente productivo para llevar a cabo investigaciones psicológicas a un nivel tan fino que pudiéramos tolerar la asignación de papeles del estado psicológico en términos sintácticos, seguiríamos queriendo tener un modo de señalar importantes *semejanzas* psicológicas entre estados opuestos en personas que piensan en diferentes lenguajes del pensamiento análogos a la semejanza entre *j'ai faim* y "tengo hambre" e *Ich habe Hunger* y entre *Es tut mir Leid* y "lo siento". Si deseamos examinar las afirmaciones que hacemos a este nivel de abstracción como afirmaciones sobre *signos del mismo tipo*, la asignación de papeles en cuestión no puede ser sintáctica (puesto que incluso en el nivel sintáctico más denso, estas expresiones son gramaticalmente distintas: tienen sustantivos donde otras tienen adjetivos, por ejemplo), ni puede ser totalmente semántico en el sentido de *proposicional*, puesto que las expresiones, con sus índices, expresan diferentes proposiciones en diferentes momentos. Lo que necesitamos es una toxonomía intermedia: los items similares serán similares en cuanto que tienen *papeles* semejantes que jugar dentro de una teoría funcionalista de los creyentes (véase Sellers, 1974).[8]

[8] Véase también Field (1978, pág. 47) quien considera afirmaciones tales como "él cree que al-

Kaplan no responde a la pregunta de si su noción de carácter puede ser aplicada interlingüísticamente. ¿Tiene *j'ai faim* exactamente el mismo carácter que "Tengo hambre"? Desearemos que nuestra contrafigura psicológica en el carácter tenga esta característica si hemos de usarla para caracterizar semejanzas psicológicas que pueden existir entre creyentes que piensan en distintos lenguajes mentalistas. Recuérdese que estamos buscando una manera de caracterizar, de la manera más general, la contribución orgánica a la fijación de actitudes proposicionales, y puesto que deseamos conceder que usted y yo podemos creer la proposición de que las ballenas son mamíferos pese a nuestras diferencias en el lenguaje mentalista, necesitamos caracterizar lo que es común en nosotros y que puede, a veces, producir la misma función de los contextos —inserciones en el mundo— a las proposiciones.

El valor de un nivel sintáctico neutral de caracterización psicológica surge más claramente cuando se considera a un ser humano —más específicamente un sistema nervioso humano— en un nivel de descripción puramente sintáctico y no interpretado. (Esto proporcionaría la materia prima, el "texto" para la interpretación semántica posterior.) Esto sería solipsismo metodológico o psicología en el sentido estrecho, puesto que estaríamos limitando nuestra visión para perder de vista todas las relaciones normales entre las cosas en el ambiente y las actividades dentro del sistema. Parte de la tarea sería distinguir el subgrupo de rasgos y regularidades físicas dentro del organismo, que anuncian rasgos y regularidades sintácticas, localizando y purificando el "texto" en medio de la confusión de garabatos y borrones. Puesto que lo que hace sintáctico a un rasgo es su capacidad de producir una diferencia semántica, esta purificación del texto no puede proseguir ignorando las presunciones semánticas, por más experimentales que sean. ¿Cómo podría funcionar esto?

Nuestro solipsismo metodológico impone que pasemos por alto el ambiente en que reside el organismo —o en el que ha residido—; sin embargo, todavía podemos localizar un límite entre el organismo y su entorno y determinar la superficie de entrada y salida de su sistema nervioso. En estas periferias hay *transductores* sensoriales y *efectores* motores. Los *transductores* responden a patrones de energía física que hacen impacto en ellos por medio de la producción de objetos sintácticos —"señales"— con ciertas propiedades. Los *efectores* en el otro extremo, responden a otros objetos sintácticos —"órdenes"— por medio de la producción de flexiones musculares de ciertos tipos. Una idea que en distinta forma autoriza toda la especulación y la teorización de la semántica de la representación mental es la idea de que las propiedades semánticas de las representaciones mentales se pueden determinar al menos parcialmente mediante sus relaciones, aunque indirectas, con estos transductores y efectores. Si conocemos las condiciones de estímulo de un transductor, por ejemplo, podemos *empezar* a interpretar su señal, sujeta a

---

gunas oraciones de su idioma que juegan aproximadamente en su psicología el papel que la oración 'hay un conejo cerca' juega en la mía". Decide que tales afirmaciones implican introducir una "noción más o menos semántica" en una teoría psicológica que se suponía liberada de los problemas semánticos. Véase también Stich, 1982, acerca de la atribución de contenido y la semejanza de éste.

muchos peligros y advertencias. Una interpretación semántica igualmente experimental y parcial de las "órdenes" puede darse una vez que veamos qué movimientos del cuerpo producen normalmente. Moviéndonos hacia el centro, alejándonos tanto de los transductores como de los efectores, podemos dotar a los sucesos y estados más centrales con poderes representativos y, en consecuencia, y por lo menos con una interpretación semántica parcial (véase, por ejemplo, Dennett, 1969, 1978a).

Por el momento, sin embargo, deberíamos cerrar los ojos a esta información de la sensibilidad del transductor y del efector y tratar a los transductores como "oráculos" cuyas fuentes de información están ocultas (y cuyas *obiter dicta* quedan por tanto sin ser interpretadas por nosotros) y tratar a los efectores como productores obedientes de efectos desconocidos. Esto podría parecer una limitación extraña del punto de vista a adoptar, pero tiene su razón de ser: es la visión de la mente por el ojo del cerebro, y es el cerebro, al final, el que realiza todo el trabajo (véase también Dennett, 1978a, capítulo 2, y 1978c). El cerebro es una *máquina sintáctica*, de manera que al fin, y en principio, las funciones de control de un sistema nervioso humano deben explicarse en este nivel o seguir siendo misteriosas para siempre. [9]

La alternativa es sostener —muy improbablemente— que el contenido o el sentido o el valor semántico podrían ser propiedades independientes, causales detectables de los hechos en el sistema nervioso. Para entender lo que quiero decir con esto, consideremos un caso más simple. Tengo dos monedas en el bolsillo y una de ellas (solamente una) *pasó exactamente diez minutos sobre mi escritorio*. Esta propiedad no es una propiedad causalmente pertinente a cómo afectará a cualquier entidad con la que entre en contacto a continuación. No hay ninguna máquina de monedas, por más compleja que sea, que pueda rechazar mi moneda probándola buscando *esa* propiedad; aunque podría rechazarla por ser radiactiva o grasienta o más caliente que la temperatura ambiente. Ahora bien, si la moneda tuviera una de estas propiedades simplemente en virtud de haber pasado exactamente diez minutos sobre mi escritorio (el escritorio es radiactivo, está cubierto de grasa, es una combinación de escritorio y horno de alfarería) la máquina de monedas se podría usar para probar en forma indirecta (y, por supuesto, no muy fiable) la propiedad de haber pasado diez minutos sobre mi escritorio. La prueba hecha por el cerebro de las propiedades *semánticas* de las señales y los estados en el sistema nervioso tiene que ser igualmente indirecta, conducida por las propiedades meramente sintácticas de los items que están siendo discriminados, es decir, por cualesquiera que sean las propiedades estructurales que los items tengan que sean aptas para un test mecánico directo. (Un test "directo" todavía no es infalible, por supuesto, o absolutamente directo.) De algún modo, el virtuosismo sintáctico de nuestro cerebro nos permite ser interpretados a otro nivel como *máquinas semánticas*: sistemas que discriminan (indirectamente) el significado de los impactos recibidos, que comprenden, significan y creen.

[9] Fodor (1980), hace hincapié en algo muy parecido a argüir en favor de lo que llama la condición de formalidad: los estados mentales pueden ser distintos (de tipo) sólo si las represntaciones que constituyen sus objetos son formalmente distintas. Véase también Field, 1978.

Esta posición ventajosa del cerebro *como máquina sintáctica* nos da un diagnóstico de lo que está ocurriendo en las argumentaciones de Putnam, Kaplan y Perry. Si el significado de algo —por ejemplo, un estado interno de almacenaje de la información, un cambio percibido en el ambiente, una elocución oída— es una propiedad detectable sólo indirectamente por un sistema tal como el cerebro de una persona, entonces el significado así concebido *no* es la propiedad a usar para armar predicados proyectables descriptivos de la conducta del sistema. Lo que queremos es más bien una propiedad que sea el significado así concebido, aproximadamente, como la propiedad de la *culpa más allá de una duda razonable es a la culpa*. Si usted quiere predecir si el jurado absolverá o condenará, la segunda propiedad es por desgracia, pero inevitablemente, un poco menos confiable que la primera.

Pero entonces podemos ver que Putnam *et al.* están metiendo una vieja cuña en una hendidura nueva: la distinción entre *real* y *aparente* está siendo adaptada para distinguir las proposiciones creídas reales y aparentes. Por tanto no es sorprendente, pero tampoco muy alentador, observar que el movimiento teórico que muchos quieren efectuar en esta situación es análogo al movimiento que con anterioridad dejó a los filósofos limitados por los datos del sentido o por las *cualidades*. Lo que es *directamente* accesible a la mente no es una característica de la superficie de las cosas *de ahí afuera*, sino una especie de copia interna que tiene vida propia. Desde esta posición de privilegio las oraciones en el lenguaje mentalista se ven como copias internas de las proposiciones que llegamos a creer en virtud de nuestra posición en el mundo. Uno no debe discutir la culpa por asociación, de manera que queda abierta la pregunta de si en este caso el movimiento teórico puede proporcionarnos un modelo útil de la mente, cualesquiera sean sus imperfecciones en aplicaciones previas. Tenemos que insistir en nuestra pregunta anterior: ¿Qué podríamos comprender acerca de un cerebro (o una mente) considerado simplemente como una máquina sintáctica?

La estrategia del solipsismo metodológico se une con el modelo de mentalidad del lenguaje del pensamiento para producir la tentadora idea de que uno podría en principio dividir la psicología en psicología sintáctica (practicada bajo el solipsismo metodológico) y la psicología semántica (que nos exigiría echarle una mirada al mundo). Hemos visto que la tarea preliminar de descubrir cuáles características internas deberían ser consideradas como sintácticas depende de suposiciones acerca de los papeles semánticos a ser jugados por los acontecimientos en el sistema, pero es tentador suponer que la sintaxis del sistema no dependerá de detalles particulares de estos papeles semánticos, sino sólo de presunciones acerca de la existencia y diferenciación de estos papeles. Supongamos, sugiere esta tentadora línea de pensamiento, que le hagamos a nuestro solipsismo metodológico el honor de desinterpretar los mensajes enviados por los transductores y las órdenes enviadas a los efectores: se considera que los transductores aseveran sólo que es F ahora, haciéndose más G y más G intermitentemente H, donde éstos son predicados sensoriales no interpretados; y los efectores obedientemente encienden el más X o el más Y o hacen que Z se mueva. ¿No seríamos entonces capaces de determinar la interpretación semántica *relativa* de estados más centrales (presu-

miblemente creencias, deseos y demás) en términos de estos predicados no interpretados? Podríamos aprender que la historia pasada del sistema lo había llevado de una manera u otra al estado de creer que todas las *F* son muy *GH*, y que el *X*-endo habitualmente conduce a una *JK* o a una *JL*. La idea de que podríamos hacer esto es paralela a la sugerencia de Field (1972) de que podríamos poner en práctica la semántica tarskiana para un lenguaje (natural) en dos partes independientemente completables: la teoría de referencia para los primitivos, y todo lo demás. Podemos hacer primero la última parte mientras contemporizamos mediante el uso de esas afirmaciones de referencia primitiva, tales como "la nieve" se refiere a lo que sea que se refiere.

La idea que capacita para la psicología de la actitud oracional es que también podríamos ser capaces de contemporizar acerca de la referencia ulterior de los predicados del lenguaje mentalista mientras seguimos rápidamente con su interpretación relativa, como lo revela una estructura semántica, sistemática y totalmente interna. Field (1978) propone esta división del problema con la ayuda de un término técnico, "creer"*.

> (1) *x* cree que *p* si y sólo si hay una oración *O* tal que *X* cree *O* y *O* significa que *p*
>
> ...el efecto de adoptar (1) es dividir el problema de dar una explicación materialísticamente adecuada de la relación de creencia en dos subproblemas:
>
> *subproblema (a):* el problema de explicar qué es para una persona creer* una oración (de su propia lengua).
>
> *subproblema (b):* el problema de explicar lo que es para una oración significar que *p*.
>
> ...La idea aproximada de cómo dar una explicación de (a) debería ser bastante clara: yo creo* una oración de mi idioma si y sólo si estoy dispuesto a emplear esa oración de cierto modo al razonar, deliberar, etcétera. Por supuesto que esto es muy vago... pero confío en que aun las vagas observaciones anteriores sean suficientes para predisponer al lector para creer que creer* no es una relación que debería ser una preocupación particular para un materialista (ni siquiera un materialista impresionado por el problema de Brentano [de la intencionalidad]). Por otra parte cualquiera que esté impresionado por el problema de Brentano *tiene posibilidades* de ser impresionado por su problema (b), puesto que a diferencia de (a), (b) invoca una relación *semántica (de significar que)* (1978, pág. 13).

De manera que se supone que creer* es una relación nada semántica entre una persona y un objeto sintácticamente caracterizado. El "cierto modo" en que uno debe estar dispuesto a emplear la creación se deja sin especificar, por supuesto, pero se debe presumir que su especificación puede completarse en principio solamente en términos sintácticos. Sólo así puede la relación no ser una "preocupación" para un materialista.

Mientras Field se quede con oraciones del lenguaje natural como las *relata* de las creencias* pisa un terreno bastante seguro al hablar de esta manera, puesto que una oración de un lenguaje natural se puede identificar independientemente de las disposiciones de cualquier persona para usarla en distintas maneras. Pero una vez que Field se vuelve hacia las oraciones del len-

guaje mentalista (o análogos de oraciones como él los llama) —como debe, por razones conocidas que tienen que ver con creyentes mudos, animales y prelingüísticos, por ejemplo— esta definición de primera aproximación de "creer"* se vuelve sumamente problemática, aunque Field mismo caracteriza el cambio como "sólo una modificación menor" (pág. 18).

Tomé sólo el caso más sencillo: "el mensaje" enviado por un elemento informador relativamente periférico cerca de la retina. Llamemos a este elemento *Inf*. Supongamos que nuestra primera hipótesis es que la señal de *Inf*, es un signo de la oración mentalista (traducida estrictamente) "ahora hay una pequeña mancha roja en el medio del campo visual". Al des-interpretar la oración vemos que tiene la forma sintáctica (para nuestros propósitos, ¿y qué otros propósitos pueden tener importancia?) *ahora hay un FGH en J de K*. Consideramos que en el mensaje están todos estos términos sólo porque suponemos que el mensaje puede contribuir en estas muchas diferentes maneras en virtud de sus vínculos. Pero quizás hayamos interpretado mal su función en el sistema. Tal vez la oración mentalista que ha de asociarse con ella (otra vez según la traducción estricta) es "Hay un tomate frente a mí" o meramente "Por lo menos diez células de la retina de la clase *F* están en el estado *G*" o quizá "Me hacen aparecer como *ruborizándome*". Estas oraciones (o por lo menos sus traducciones, como se ha visto —tienen análisis sintácticos completamente distintos, ¿pero qué forma sintáctica tiene *esa cosa* que hemos localizado en el cerebro? Podemos ser capaces de determinar la "forma" de un ítem —un tipo de suceso, por ejemplo— en el cerebro, pero no podemos determinar su forma sintáctica (diferenciada de sus propiedades meramente decorativas, por más distintivas que sean, excepto determinando sus poderes especiales para combinarse o cooperar con los otros elementos, y últimamente su importancia *ambiental* por la vía de esos poderes de interacción.

Un experimento del pensamiento presentará el punto con más claridad. Supongamos que nuestra tarea fuera diseñar un lenguaje de pensamiento más que descubrir uno ya en funcionamiento. Resolvemos lo que queremos que nuestro sistema crea (desee, etc.) y anotamos versiones de toda esta información en oraciones de alguna variedad tentativa del lenguaje mentalista. Inscribimos cada oración de creencia en un compartimento separado de un gran cuadro del sistema que estamos diseñando. Un compartimento de creencia tiene la traducción en lenguaje mentalista de "la nieve es blanca". El simple hecho de tener los símbolos escritos en el compartimento no puede almacenar la información de que la nieve es blanca, por supuesto. Por lo menos debe haber algún mecanismo preparado para utilizar los símbolos de este compartimento de manera tal que establezca una diferencia, el tipo exacto de diferencia, para creer que la nieve es blanca. Este mecanismo debe, por ejemplo, unir de algún modo el compartimento que contiene "la nieve es blanca" con todos los compartimentos en los cuales hay oraciones con la palabra en lenguaje mentalista que quiere decir "blanco". Estos compartimentos están unidos entre sí en ciertas formas sistemáticas y finalmente, de algún modo, con la periferia del sistema, con el mecanismo que podría señalar la presencia de materia blanca fría, materia verde fría... El compartimento

"la nieve es blanca" también está unido a todos los compartimentos con "nieve", que a su vez están unidos a todos los compartimentos con "precipitaciones", etcétera. La vasta red imaginaria de nexos convertiría a la colección de compartimentos en *algo parecido* a los enrejados taxonómicos o redes hereditarias estructurales o redes semánticas que se encuentran en los sistemas de Inteligencia Artificial (véase, por ejemplo, Woods, 1975, 1981). Además están todos los nexos con el mecanismo, cualquiera que sea, que confía en estos compartimentos para contribuir apropiadamente al control del comportamiento de todo el sistema, o mejor aun, del ser en el cual está inserto el sistema. Sin todos estos nexos, las inscripciones en los compartimentos son simple ornamento: no almacenan la información sean como fueren. Pero del mismo modo, una vez que los nexos están en su lugar, las inscripciones en los compartimentos siguen siendo simple adorno, o en el mejor de los casos, rótulos mnemotécnicos encapsulando *para nosotros* (más o menos exactamente) la información verdaderamente almacenada *para el sistema* en ese nodo, en virtud de los nexos desde el nodo a otros nodos del sistema. La verdadera "sintaxis", la estructura del sistema de la cual depende la función está toda en los nexos. (Estoy usando "nexos" como comodines para lo que haga falta para jugar este papel; nadie sabe todavía (que yo sepa) cómo resolver este problema en concreto.)

La separación imaginada entre los nexos y las inscripciones en los compartimentos de nuestro ejemplo no refleja, por supuesto, la situación real en la Inteligencia Artificial. La cuestión de los lenguajes de computación es que están diseñados con mucha inteligencia de manera que sus instrucciones adecuadamente ingresadas en el sistema, *crean* distintos nexos con elementos en otras inscripciones. Es esta característica la que hace a los lenguajes de la computación tan diferentes de los lenguajes naturales, y con seguridad es una característica que cualquier lenguaje del pensamiento tendría que tener, si su postulado es evitar el epifenomenalismo improductivo o nuestros rótulos imaginarios en los compartimentos. Y para cualquier "lenguaje" que tenga esta característica, la relación entre forma y función es sin duda estrecha, tan estrecha que la distinción en el esquema ampliado de Kaplan en la etapa (0) entre la contribución de los rasgos sintácticos y la contribución de las "convenciones lingüísticas" no tiene ninguna contrafigura en ningún modelo de "lenguaje del pensamiento" plausible para la psicología. Obtener el "texto" independientemente de obtener su interpretación no es una perspectiva real para la psicología. De manera que la división propuesta por Field del problema de la creencia en el subproblema *(a)*, el problema sintáctico, y el subproblema *(b)*, el problema semántico, tomada como una propuesta para una estrategia de investigación es vacía.

No obstante, tal vez la propuesta de Field se pueda reconstruir, no como una recomendación para un programa de investigación sino para marcar una diferencia muy importante del razonamiento. Aun cuando no podamos en realidad determinar *primero* la sintaxis y *luego* la semántica de un lenguaje del pensamiento (por razones epistemológicas), la diferencia podría indicar todavía algo real, de manera que habiéndonos abierto con gran esfuerzo el camino tanto hacia una teoría semántica como sintáctica del lenguaje

del pensamiento pudiéramos distinguir las propiedades sintácticas del sistema de sus propiedades semánticas. Sin duda, una distinción parecida a la distinción entre semántica y sintaxis se podrá hacer en retrospectiva en cualquier psicología de la creencia madura y confirmada, puesto que tiene que haber alguna manera de describir el funcionamiento del sistema nervioso con independencia de su inserción en el mundo, en virtud de la cual fijamos su caracterización semántica. Pero suponer que esta distinción tuviera mucho en común con la distinción entre la sintaxis y la semántica para un lenguaje natural, es comprometernos gratuitamente con un fuerte oracionismo.

Se puede extraer una moraleja más fuerte de la discusión del problema de asociar un *mensaje explícito* con la contribución de Inf, el transductor visual periférico. Supusimos que habíamos aislado a Inf. como componente funcional del sistema cognitivo, un componente al que se le veía *informando* al sistema acerca de algún rasgo visual. Enfrentamos entonces el problema de aparecer con una *lingüificación* apta de esa contribución, es decir encontrando una oración que exprese de manera explícita y exacta el mensaje aseverado por Inf. Y vimos que la oración que eligiéramos dependía de manera crítica, exactamente, de los poderes combinatorios que tenía realmente el mensaje de Inf. Aun suponiendo que pudiéramos determinar ésto, aun suponiendo que pudiéramos demostrar que teníamos una mejor descripción funcional del sistema del cual Inf es una parte, y pudiéramos decir *exactamente* qué funciones puede desempeñar la señal de Inf, no hay ninguna garantía de que esas funciones sean consultadas o aludidas de manera apta y exacta por ninguna afirmación en el sentido de que el mensaje de Inf inserta la oración *O* (en el lenguaje *L*) como premisa en algún sistema deductivo o de inferencia. Me parece que la convicción de que debe ser posible lingüificar cualquier contribución así subyace bajo una gran cantidad de ideología metateórica reciente, y sospecho que se apoya en parte en una fusión equivocada *de determinación* y *precisión*.[10] Supongamos que la contribución semántica de Inf, su *informe* del sistema, sea enteramente definido. Es decir, podemos decir exactamente como el hecho de que ocurra produciría efectos que se ramifican a través del sistema produciendo diferencias en el contenido o en la contribución semántica de otros subsistemas. Sin embargo, no significaría que pudiéramos hacer *explícita* esta contribución en forma de una oración u oraciones válidas. Tendríamos que tener alguna manera de *describir* la contribución semántica de manera perfectamente explícita, sin describirla como una afirmación explícita en algún idioma.

No aludo sólo a la posibilidad de que la actividad en una parte de un sistema cognitivo pudiera tener un efecto sistemático aunque *ruidoso* —o en todo caso negativo— en la otra parte. Esos efectos son muy posibles. El olor a azufre podría hacer que alguien pensara en el béisbol *sin ningún motivo*. Es decir, no podría haber un nexo *informativo* tal como el recuerdo de partidos olorosos jugados detrás de una fábrica de gas, pero sin embargo sería un nexo *causal* fiable, si bien sin sentido. Admito la posibilidad de dichos efectos pero estoy abogando por algo más fuerte: que pudieran haber relaciones diseña-

[10] Las discusiones de Charles Taylor acerca de la precisión y de la "explicitación" ayudaron a dar forma a las afirmaciones del resto de esta sección.

das entre las actividades de los distintos subsistemas cognitivos que fueran de alto contenido sensible, informativo, epistémicamente útiles y que a su vez desafiaran, sin embargo, la interpretación oracionista. Supongamos, por ejemplo, que Pat dice que Mike "tien algo contra los pelirrojos". Lo que Pat quiere decir es, aproximadamente, que Mike tiene un estereotipo de un pelirrojo que es más bien despectivo y que influye en las expectativas de Mike acerca de los pelirrojos y en sus interacciones con ellos. No se trata simplemente de que él tenga un prejuicio contra los pelirrojos, sino de que tiene algo idiosincrático y *particular* hacia ellos. Y Pat podría tener razón, ¡más razón de la que él mismo cree! Podría resultar que Mike realmente tiene algo, un algo de mecanismo cognitivo, *acerca de los pelirrojos* en el sentido de que sistemáticamente entra en acción cada vez que el tema son los pelirrojos y que adapta distintos parámetros del mecanismo cognitivo formulando hipótesis halagadoras acerca de los pelirrojos muy poco probables de ser tenidas en cuenta o confirmadas, haciendo que la conducta relativamente agresiva *vis-à-vis* los pelirrojos más cerca de la realización de lo que estaría de otro modo, etcétera. Ese algo *contra los pelirrojos* podría ser de funcionamiento muy complejo o muy simple y en cualquiera de los dos casos su papel eludiría la caracterización en el formato:

Mike cree que: (x) (x es un pelirrojo $\supset$ ...)

independientemente de lo tortuosamente que amontonáramos las cláusulas de exclusión, los calificadores, los operadores de probabilidades y otros adaptadores explícitos del contenido. La contribución de ese algo de Mike acerca de los pelirrojos podría ser perfectamente definido y también innegablemente lleno de contenido y sin embargo ninguna lingüificación de él podría ser más que un rótulo mnemotécnico para su papel. En tal caso podríamos decir, como con frecuencia hay razones para hacerlo, que hay distintas creencias *implícitas* en el sistema. Por ejemplo, la creencia de que los pelirrojos no son dignos de confianza. ¿O tendría que ser la creencia de que la mayoría de los pelirrojos no son dignos de confianza o "todos los pelirrojos que he conocido"? ¿O tendría que ser "(x) (x es un pelirrojo $\supset$ la probabilidad es de 0,9 de que x no sea digno de confianza)"? La preocupación sobre la forma adecuada de la oración es ociosa cuando la oración es sólo parte de un esfuerzo por captar el contenido implícito de alguna parte del mecanismo que no es oracional. (Véase "Las actitudes proposicionales" más atrás acerca de los debates fútiles ocasionados por el no poder distinguir las proposiciones de las oraciones.)

Todavía se podría argumentar que aunque los contribuyentes semánticos *no* oracionales del tipo que acabo de esbozar pudieran jugar un gran papel en el mecanismo cognitivo del cerebro humano, hacen falta también estados representativos explícitamente oracionales, aunque más no sea para darle al cerebro limitado la composicionalidad que necesita para representar infinitamente muchos diferentes estados de cosas con sus recursos limitados. Según lo que puedo ver, esto sería posible. Por cierto que algún tipo de *composicionalidad* muy eficiente y elegante explica los poderes esencialmente ilimitados que tenemos para percibir, meditar, creer, intentar... distintas co-

sas. Los únicos ejemplos que ahora tenemos de sistemas *universales* (discutibles) de representación con medios limitados son los lenguajes, y tal vez cualquier sistema de representación universal posible debe ser reconocidamente oracional, en un sentido aún por determinar, por supuesto. Supongamos que esto es así (y siéntanse en buena compañía). Por tanto, aunque una gran parte de la psicología pudiera ser tanto cognitiva como *no* oracional, en el núcleo de la persona estaría su sistema oracional. La teoría de la percepción visual, por ejemplo, podría requerir el lenguaje mentalés sólo en alguna "interfaz" relativamente central con respecto al núcleo oracional. Al principio considerábamos que la tarea de la psicología de la actitud oracional empezaba en los transductores y efectores periféricos y se abría camino hacia adentro por medio de oraciones. Quizás el error consistiera en suponer que el interfaz entre el sistema oracional y el *mundo* —el abrigo de los transductores y efectores usado por todos los sistemas cognitivos— era más liviano de lo que es.

Grueso o liviano, el abrigo de los mecanismos transductor-perceptor y los mecanismos efector-actor se convierte en una especie de ambiente, un contexto, en el cual tienen lugar las "elocuciones", postuladas en el lenguaje mentalés. Al ignorar ese contexto, los predicados del lenguaje mentalés quedan *muy* sin interpretar; aun tomando en cuenta ese contexto, los predicados del lenguaje mentalés serían interpretados sólo parcialmente —no serían interpretados de manera lo bastante completa, por ejemplo, como para distinguir mis actitudes proposicionales de las de mi *Doppelgänger*. Podemos armar algunas variaciones extrañas del experimento de Putnam utilizando esta idea. Supongamos que un duplicado físico de esa parte de mi sistema nervioso que es el sistema oracional quedara enganchado (aquí mismo en la Tierra) con un abrigo diferente de transductores y efectores. Mi sistema oracional puede almacenar la información de que todas las *F* son muy *GHs*, y lo mismo puede hacer su réplica, pero en mí, el estado oracional apoya[11] mi creencia de que todos los galgos son animales muy veloces, mientras que en el otro ser apoya la creencia de que los palacios son casas muy caras. Estos signos diferentes del mismo tipo *sintáctico* tienen el mismo *carácter* si tratamos el abrigo transductor-efector como parte indistinta del contexto "externo" de la elocución; si trazamos un límite más lejano entre el abrigo y el ambiente, entonces estos signos no tienen el mismo carácter, sino sólo la misma sintaxis, y los diferentes abrigos juegan los papeles de contrafiguras de las convenciones lingüísticas de la etapa (0) de Kaplan.

El punto es que el esquema de Kaplan es un caso especial de algo muy general. Cada vez que describimos un sistema funcional, si trazamos un límite entre el sistema "propiamente dicho" y algún contexto o nicho ambiental en el que reside, descubrimos que podemos caracterizar un esquema al estilo de Kaplan.

$$C + E \rightarrow I$$

[11] "Apoya" es un término gestual por el cual podemos agradecer a los neurofisiólogos. Si reunimos dos ejemplos de una jerga, podemos decir que una creencia super sigue al estado que la *apoya*.

donde C es un concepto semejante al carácter de aplicación estrecha o intrasistémica; E es el concepto de un contexto de inserción o ambiente de funcionamiento, e I es una caracterización semántica (o funcional) *más rica* del papel sistémico en cuestión que la proporcionada por C solo. Donde el sistema en cuestión es un sistema representativo o de creencia, "más rico" significa más próximo a determinar una proposición (clásica), o si incluimos la etapa (2) de Kaplan como el paso final de esta progresión, más rico en el sentido de estar más cerca de la referencia final a las cosas del mundo. En otros contextos —tales como las caracterizaciones de los componentes funcionales en biología o ingeniería (véase Wimsalt, 1974)— la caracterización "más rica" nos dice más del punto funcional del ítem: lo que se ve limitadamente como un productor de chispa resulta ser en contexto un deflagrador de combustible, para tomar un ejemplo exagerado.

Pasando de una etapa a otra en semejante esquema de interpretación, se ve que cuanto más rica es la semántica de una etapa determinada, más abstracta o tolerante es la sintaxis. Las oraciones con diferentes propiedades físicas pueden tener la misma sintaxis. Las oraciones con distinta sintaxis pueden tener el mismo carácter. Las oraciones con distinto carácter pueden expresar la misma proposición. Finalmente, proposiciones diferentes pueden atribuirle la misma propiedad al mismo individuo: el hecho de que el Director de Admisiones sea de mediana edad no es idéntico a la proposición de que el Director más alto es de mediana edad. Transplantada de la teoría del lenguaje natural a la teoría de los estados psicológicos, la parte donde anida el interés la vemos ahora con este aspecto: la gente que cree la misma proposición puede estar en estados psicológicos distintos (restringidos); la gente que está en un mismo estado psicológico restringido puede estar en distintos estados psicológicos de grano fino (es decir, caracterizados sintácticamente; las personas que están en los mismos estados sintácticos pueden implementar esos estados de maneras físicamente diferentes. Y, por supuesto, al mirar hacia la otra dirección, vemos que dos personas limitadamente interpretadas como estando en el mismo estado pueden ser reinterpretadas como estando en estados diferentes si volvemos a trazar los límites entre los estados de las personas y el ambiente que las rodea.[12]

Actitudes nocionales

Frente a las objeciones de Putnam y otros hacia las actitudes proposicionales "clásicas", advertimos acerca de la pregunta: ¿cuál es la contribución orgánica en la determinación de las actitudes proposicionales? La respuesta caracterizaría a los estados psicológicos "en sentido restringido". La tentativa de captar estos tipos de estado psicológico restringido como actitu-

[12] Burge (1979) presenta un extenso experimento del pensamiento sobre las creencias acerca de la artritis que puede ser considerado como el trazado del límite entre el sistema propiamente dicho y su entorno completamente *afuera* del individuo biológico. Las variaciones contextuales abarcan prácticas *sociales* fuera de la experiencia del sujeto. (Para una crítica de Burge, véase el cap. 8).

des racionales chocó con una variedad de problemas, el principal de los cuales fue que cualquier caracterización de actitud oracional, al ser esencialmente una asignación sintáctica de papeles, dejaría poco margen de error. En el experimento de Putnam admitimos que la duplicación *física* es suficiente pero no necesaria para la identidad de la contribución orgánica; también podríamos conceder que la semejanza más débil captada por la duplicación *sintáctica* (en cierto nivel de abstracción) sería suficiente para la identidad de la contribución orgánica, pero aun cuando la identidad de la contribución orgánica —la capacidad gemela psicológica restringida— sea una condición muy rigurosa, no parecería necesitar de la capacidad gemela sintáctica a ningún nivel de descripción. Considérese la pregunta de alguna forma análoga: ¿todas las máquinas de Turing que computan la misma función comparten (es decir, mesa de la máquina) una descripción sintáctica? No, a menos que adaptemos nuestros niveles de descripción de la mesa de la máquina y el comportamiento de entrada y salida de datos como para que se unan de manera trivial. Qué es lo que debería tomarse en cuenta como una equivalencia de las máquinas de Turing (o programas de las computadoras) es una pregunta fastidiosa; no lo sería si no fuera por el hecho de que descripciones diferentes no triviales en términos de "sintaxis" interna pueden dar como resultado la misma "contribución" en algún nivel de descripción útil.

No hay duda de que la analogía es imperfecta y que otras consideraciones —por ejemplo, consideraciones biológicas— podrían pesar a favor de la suposición de que la capacidad gemela psicológicamente restringida *completa*, necesita la capacidad gemela sintáctica en algún nivel, pero incluso si se admitiera esto, no significaría que la semejanza psicológica parcial se pueda describir siempre en algún sistema general de descripción aplicable a todos los que comparten el rasgo psicológico. Las personas vanidosas o paranoicas, por ejemplo, son seguramente parecidas psicológicamente; una gran parte de la semejanza en cada caso parecería estar bien captada cuando se habla de creencias semejantes o compartidas. Aun si se adopta una línea rigurosa contraproducente acerca de la identidad de la creencia (según la cual no hay ni dos personas que alguna vez compartan de verdad una creencia), estas *semejanzas* en la creencia piden ser captadas dentro de la psicología. No se podría sostener de manera creíble que dependan del monolingüismo los cerebros de personas vanidosas que hablan todos el mismo lenguaje mental. Tampoco podemos captar estas semejanzas en el estado de creencia por la vía de las actitudes *proposicionales*, debido a la indicatividad de muchas de las creencias cruciales: "La gente *me* admira", "La gente quiere destruir*me*".

Estas consideraciones sugieren que lo que estamos buscando caracterizar es una posición intermedia, a mitad de camino entre la semántica y la sintaxis, se podría decir. Llamémosla *psicología de la actitud nocional*. Queremos que resulte que yo y mi *Doppelgänger* —y cualquier otro par de gemelos psicológicamente restringidos— tenemos exactamente las mismas actitudes nocionales, de manera que nuestras diferencias en las actitudes proposicionales se deben enteramente a las diferentes contribuciones ambientales. Pero también queremos que resulte que usted y yo, que no somos gemelos psicológicos "pero sí de mentes parecidas" acerca de varios tópicos, compartimos una diversidad de actitudes nocionales.

Una idea familiar que ha aparecido con muchos aspectos se puede adaptar a nuestros propósitos aquí: la idea del mundo subjetivo de una persona "El mundo en que vivo" de Hellen Keller o "El mundo según Garp", de John Irving, por ejemplo. Tratamos de caracterizar el *mundo nocional* de un sujeto psicológico de manera tal que, por ejemplo aunque mi *Doppelgänger* y yo vivamos en mundos reales diferentes —el Planeta Tierra Gemelo y la Tierra— tenemos *el mismo* mundo nocional. Usted y yo vivimos en el mismo mundo real pero tenemos diferentes mundos nocionales, aunque hay una considerable superposición entre ellos.

Un mundo nocional debería ser considerado como una especie de mundo de *ficción* ideado por un teórico, un tercer observador, para caracterizar los estados psicológicos restringidos de un individuo. Se puede suponer que un mundo nocional está lleno de objetos nocionales y el escenario de sucesos nocionales —todos los objetos y sucesos en los que el sujeto cree, se podría decir—. Si relajamos un momento nuestro solipsismo metodológico notaremos que algunos objetos del mundo real habitado por un sujeto "hacen juego" con los objetos del mundo nocional del sujeto, pero otros no. El mundo real contiene muchas cosas y sucesos que no tienen contrafigura en el mundo nocional de ningún sujeto (excluyendo el mundo nocional de un Dios omnisciente) y los mundos nocionales de los sujetos crédulos o confundidos u ontológicamente licenciosos contendrán objetos nocionales que no tienen contrafiguras en el mundo real. La tarea de describir las relaciones que pueden existir entre las cosas del mundo real y las cosas del mundo nocional de alguien está notoriamente plagada de enigmas. Esa es una razón por la cual retirarse al solipsismo metodológico: descomponer temporalmente en factores esos puntos conflictivos.

Nuestra retirada nos ha llevado a un terreno muy conocido: ¿qué son los objetos nocionales sino los *objetos intencionales* de Brentano? El solipsismo metodológico es aparentemente una versión de la *époché*, o clasificación, de Husserl. ¿Puede ser que la alternativa tanto para la psicología de la actitud proposicional como para la psicología de la actitud oracional sea... la Fenomenología? No del todo. Hay una diferencia esencial entre el enfoque a esbozar aquí y los enfoques tradicionales asociados con la fenomenología. Allí donde los fenomenologistas proponen que uno puede llegar a su *propio* mundo nocional por medio de una gimnasia mental introspectiva algo especial —llamada por algunos la reducción fenomenológica— a nosotros nos interesa determinar el mundo nocional de *otro* desde afuera. La tradición de Brentano y Husserl es la *autofenomenología;* lo que yo propongo es la *heterofenomenología* (véase "Two Approaches to Mental Images" en *Brainstorms*). Aunque los resultados pudieran tener una semejanza notable, las suposiciones habilitantes son muy distintas.

Se puede ver mejor la diferencia con la ayuda de una distinción recientemente devuelta a la circulación por Fodor (1980) entre lo que él llama, siguiendo a James, psicología *naturalista* y *racional*. Fodor cita a James:

> En general pocas fórmulas recientes le han prestado un mejor servicio de un tipo aproximado a la psicología que la *spenceriana*, que afirma que la esencia de la vida mental y de la vida corporal es una, es decir, "la adaptación de las rela-

> ciones internas a las externas". Semejante fórmula es la vaguedad personificada, pero puesto que toma en cuenta el hecho de que las mentes habitan en ambientes que actúan sobre ellas y respecto de los cuales ellas reaccionan a su vez; debido, en pocas palabras, porque toma a la mente en medio de todas sus relaciones concretas, es inmensamente más productiva que la anticuada "psicología racional" que trataba al alma como una existencia aparte, autosuficiente, cuya naturaleza y propiedades se daba por sentada (James, 1890, pág. 6).

James le canta loas a la psicología naturalista, la psicología en sentido *amplio* pero la moraleja del Planeta Gemelo de la Tierra, extraída explícitamente por Fodor, es que la psicología naturalista arroja sus redes demasiado lejos para que sea factible. Los fenomenologistas extraen aparentemente la misma conclusión y ambos se vuelven hacia las distintas versiones del solipsismo metodológico: interés en el sujeto psicológico "como una existencia aparte, autosuficiente", pero cuando "consideran su naturaleza y sus propiedades", ¿qué encuentran? Los fenomenologistas, utilizando cierto tipo de introspección alegan haber encontrado una experiencia dada que se convierte en la materia prima para su construcción de sus mundos nocionales. Si Fodor, usando cierto tipo de inspección interna (imaginada) del mecanismo, afirmó haber encontrado un texto en lenguaje mentalés *dado* en el hardware (que se convertiría en la materia prima para la construcción de las actitudes nocionales del sujeto) tendríamos las mismas razones para dudar de la existencia de lo dado en este caso como en el caso de la fenomenología (véase capítulo 8). James tiene razón: no se puede hacer *psicología* (como opuesta a, digamos, la neurofisiología) sin determinar las propiedades *semánticas* de los hechos y estructuras internas que se están estudiando, y no se pueden descubrir las propiedades semánticas sin observar las relaciones de esos hechos o estructuras internas con las cosas del ambiente del sujeto. No está escrito en ninguna parte que el ambiente en relación con el cual determinamos las propiedades semánticas de un sistema así, deba ser un ambiente *verdadero*, o el ambiente *real* en el que ha crecido el sistema. Un ambiente imaginario idealizado o ficticio, servirá lo mismo. La idea es que para poner en práctica la teoría de la "representación mental" hay que poner en práctica la semántica de las representaciones desde el principio. No se puede hacer primero la sintaxis y luego la semántica. Pero eso significa que hace falta un modelo, en sentido de la semántica tarskiana. Un modelo ficticio, sin embargo, permitirá la puesta en marcha de la suficiente semántica tarskiana como para que determinemos la semántica parcial o protosemántica que necesitamos para caracterizar la contribución orgánica.

La idea de un mundo nocional es, entonces, la idea de un modelo —pero no necesariamente el modelo real, verdadero, existente— de nuestras representaciones internas. *No consiste en representaciones sino en representados.* Es el mundo "en que vivimos", no el mundo de las representaciones que están dentro de mí. (Hasta ahora, esto es Brentano puro, por lo menos como lo entiendo yo. Véase Aquila, 1977). El teórico que quiera caracterizar los estados psicológicos restringidos de un ser o, en otras palabras, la contribución orgánica de ese ser a sus actitudes proposicionales, describe un mundo de ficción; la descripción está en el papel, el mundo de ficción no existe,

pero los habitantes del mundo ficticio son tratados como referentes nocionales de las representaciones del sujeto, como los objetos intencionales de ese sujeto. Se espera que, mediante el uso de esta táctica, el teórico pueda obtener las ventajas del naturalismo de James y de Spencer sin las dificultades planteadas por Putnam y los demás.

Ahora la pregunta es: ¿qué guía nuestra construcción del mundo nocional de un organismo? Para dramatizar el problema, supongamos que recibimos una caja que contiene un organismo vivo de no sabemos dónde, pero: congelado (o comatoso), y por tanto aislado de cualquier ambiente. Tenemos una instantánea laplaceana del organismo —una descripción completa de su estructura y composición internas— y podemos suponer que esto nos permite determinar exactamente cómo *respondería* a cualquier impacto ambiental nuevo en caso de que lo liberáramos de su estado de vitalidad suspendida y de su aislamiento. Nuestra tarea es como el problema que se presenta cuando nos muestran algún artefacto extraño o antiguo y nos preguntan: "¿Para qué sirve? ¿Es un aparato para fabricar agujas o un aparato de medir la altura de los objetos distantes o un arma? ¿Qué podemos aprender si estudiamos ese objeto? Podemos determinar cómo encajan las partes, qué pasa en distintas condiciones... También podemos buscar cicatrices y abolladuras reveladoras, signos de desgaste. Una vez que hemos recopilado todos esos datos, tratamos de imaginarnos una escena en la que, una vez en posesión de estos datos pudieran cumplir *de manera excelente* alguna función imaginariamente útil. Si el objeto fuera un zurcidor de velas o un deshuesador de cerezas tan bueno en una como en otra tarea no podemos descubrir qué es *en realidad* —para qué sirve— sin saber de dónde vino, quién lo fabricó y por qué. Estos datos podrían haber desaparecido sin dejar rastro. No sería completamente imposible, entonces, determinar la identidad de ese objeto o su esencia por más concienzudamente, que estudiásemos el objeto. eso no significaría que no existiera el hecho acerca de si la cosa era un deshuesador de cerezas o un zurcidor de velas, sino que la verdad, cualquiera que fuera, ya no importaba. Sería uno de esos hechos históricos vanos o inertes, como el hecho de que parte del oro que tengo en los dientes alguna vez perteneció —o no— a Julio César.

Una vez que estamos frente a nuestro novel organismo podemos determinar con bastante facilidad para qué sirve —es para sobrevivir y prosperar y reproducir su especie— y deberíamos tener pocos problemas para identificar sus órganos de los sentidos, sus modos de acción y sus necesidades biológicas. Puesto que *ex hypothesi* podemos suponer lo que haría si... (para llenar todos los blancos del antecedente), podemos determinar, por ejemplo, que comerá manzanas pero no pescado, que tiende a evitar los lugares muy iluminados, que está dispuesto a producir ciertos ruidos en ciertas condiciones, etcétera. Ahora bien, ¿para qué clase de ambiente lo harían apto estos talentos e inclinaciones? Cuanto más aprendemos de la estructura interna, las tendencias de conducta y necesidades sistémicas del organismo, más especial se vuelve nuestro ambiente ideal hipotético. Por "ambiente ideal" no quiero decir el mejor de los mundos para este organismo ("donde brota la limonada y cantan los pájaros azules de la felicidad"), sino el ambiente (o tipo de ambientes —para lo que el organismo como está corrientemente constituido—, es más apto. Sería un mundo francamente detestable, pero al menos el orga-

nismo está preparado para arreglárselas con lo que tiene de detestable. Podemos aprender algo de los enemigos del organismo —reales o sólo nocionales— si observamos su coloración productora o su conducta para escapar o... cómo contestaría ciertas preguntas.

Mientras que el organismo del que nos estamos ocupando sea muy simple y tenga, por ejemplo, poca o ninguna plasticidad en el sistema nervioso (de manera que no puede aprender), el límite de especificidad para el ambiente ideal imaginado puede no distinguir los ambientes radicalmente diferentes aunque igualmente bien equipados, como en el caso del artefacto. A medida que aumenta la capacidad para aprender y recordar, y a medida que crece la riqueza y complejidad de las relaciones posibles con las condiciones ambientales (véase el capítulo 2), la clase de modelos igualmente aceptables (ambientes ideales hipotéticos) se reduce. Más aun, en los seres que tienen la capacidad de aprender y de almacenar información de su mundo en la memoria, entra a jugar un principio exegético nuevo y más poderoso. Las cicatrices y abolladuras del deshuesador de cerezas (¿opera un zurcidor de velas?) pueden en ocasiones resultar delatoras, pero las cicatrices y huellas en la memoria de un ser que aprende están *diseñadas* para ser delatoras, para grabar con alta fidelidad tanto los encuentros especiales como las lecciones generales, para su uso en el futuro. Puesto que las cicatrices y huellas de la memoria están destinadas a ser usadas en el futuro, podemos tener esperanzas de "leerlas" utilizando nuestro conocimiento de las disposiciones que de ellas dependen, siempre que demos por sentado que las disposiciones así vinculadas son adecuadas en general. Esas interpretaciones de los "rastros de la memoria" producen información más específica del mundo en el que vivió el ser y al que se había adaptado. Pero podremos distinguir la información de este mundo de la información errónea, y entonces, la información que extrapolemos como *constituida* por el estado actual del organismo será un mundo ideal, no en el sentido de óptimo, sino en el sentido de *irreal*.

Los naturalistas insistirán, con toda razón, en que el ambiente verdadero, tal como se lo encontró, ha dejado su huella en el organismo y le ha dado una forma intrincada; el organismo está en su estado actual *debido a su historia*, y sólo esa historia podría haberlo llevado a su estado presente. Pero si estamos en un estado de ánimo dispuesto para la experimentación del pensamiento podemos imaginarnos creando un duplicado cuya historia *aparente* no fuera su verdadera historia (como en el caso de una antigüedad falsificada, con sus indicios simulados de "sufrimiento" y deterioro). Un duplicado así completo (que sólo es posible de una manera lógica, experimental) es el caso límite de algo real y conocido: cualquier rasgo particular del estado actual puede ser espurio, de modo de que la manera en que el mundo tendría que haber sido para que el ser se encuentre ahora en este estado no es exactamente como el mundo era. El mundo nocional que describimos por extrapolación del estado actual no es entonces exactamente el mundo que suponemos creó dicho estado, incluso si conocemos ese mundo real, sino más bien el mundo aparente del ser, el mundo aparente *para* el ser tal como se manifiesta en la disposición total y actual de éste.

Supóngase que apliquemos este ejercicio imaginario en la formación del mundo nocional a organismos altamente adaptables como nosotros mismos.

Esos organismos tienen una estructura interna y características naturales tan ricas en información sobre el ambiente en el que crecieron que en principio podríamos decir: este organismo está mejor equipado para un ambiente en el que hay una ciudad llamada Boston, en la que el organismo pasó su juventud, en compañía de organismos llamados... etcétera. No podíamos distinguir Boston de la Boston del Planeta Tierra Gemelo, por supuesto, pero excepto por esas variaciones virtualmente imperceptibles sobre un tema, nuestro ejercicio acerca de la formación del mundo nocional terminaría en una solución única en su género.

De todos modos ése es el mito. Es un mito prácticamente inútil, por supuesto, pero teóricamente importante, puesto que revela las presunciones fundamentales que se hacen de la dependencia final de la contribución orgánica a la constitución física del organismo. [Esta dependencia se conoce de otra manera como la supervivencia de los rasgos psicológicos (restringidos) sobre los rasgos físicos; véase, por ej., Stich, 1978a.] Al mismo tiempo, el mito preserva la subdeterminación de la referencia final que era la moraleja proclamada de las consideraciones putnamianas. Si hay un lenguaje del pensamiento ésta es la manera en que habría que abrirse camino, con esfuerzo, para descubrirlo y traducirlo sin siquiera el beneficio de tener intérpretes bilingües o pruebas circunstanciales acerca del origen del texto. Si hay alguna alternativa del tercero excluido para el dudoso método introspeccionista (*auténticamente* solipsístico) de los fenomenólogos, en el caso de que la heterofenomenología sea de alguna manera posible, tendrá que ser mediante este método.

En principio, entonces, los frutos finales del método aplicados a un ser humano bajo la compulsión del solipsismo metodológico serían una descripción exhaustiva del mundo nocional de esa persona, completo con sus identidades equivocadas, sus quimeras y sus duendes personales, sus errores y distorsiones reales.[13] Podemos pensar que es *el* mundo nocional del individuo, pero por supuesto la descripción más exhaustiva posible no lograría especificar un mundo único. Por ejemplo, las variaciones en un mundo totalmente más allá del alcance de la comprensión o intereses de una persona generaría diferentes mundos posibles igualmente compatibles con la determinación máxima dada por la constitución de la persona.

[13] ¿Y qué hay de los objetos de sus temores, esperanzas y deseos? ¿Son ellos habitantes del mundo nocional del sujeto, o debemos agregar un mundo de deseo, un mundo de temor y así sucesivamente al mundo de creencias del sujeto? (Joe Camp y otros han recalcado esta preocupación.) Cuando algo que el sujeto cree que existe es también temido o deseado por él, no hay problema: algún habitante de su mundo nocional simplemente está coloreado de deseo o temor o admiración o lo que sea. Cómo tratar "la casa de ensueño que espero construir algún día" es otra cosa. Postergando los detalles para otra ocasión, aventuraré algunas imprudentes afirmaciones generales. Mi casa de ensueño no es un habitante de mi mundo nocional a la par de mi casa real o siquiera de la casa en que terminaré mis días. Pensar en *ella* (mi casa de ensueño) no ha de ser analizado, por ejemplo, del mismo modo que pensar en mi casa o en la casa en que terminaré mis días. (Hay más sobre este tema en la sección siguiente.) Mi casa de ensueño se constituye indirectamente en mi mundo nocional por la vía de lo que podríamos llamar mis *especificaciones*, que son habitantes perfectamente comunes de mi mundo nocional y de mis creencias generales y otras actitudes. Yo creo en mis especificaciones, que ya existen en el mundo como ítems de los pertrechos mentales creados por mi pensamiento, y entonces hay creencias y deseos generales y

La situación es semejante a aquella de los mundos ficticios más conocidos tales como el mundo de Sherlock Holmes o el Londres de Dickens. Lewis (1978) da una explicación de la "verdad en la ficción", la semántica de la interpretación de la ficción que desarrolla la idea que necesitamos: "el" mundo de Sherlock Holmes se concibe formalmente mejor como un *conjunto* de mundos posibles, aproximadamente: todos los mundos posibles compatibles con la totalidad de los textos de Sherlock Holmes escritos por Conan Doyle.[14] Del mismo modo, "el" mundo nocional que describimos podría ser visto formalmente como el conjunto de mundos posibles compatibles con la descripción máxima (véanse Hintikka, 1962; Stalnaker, 1984). Obsérvese que la descripción es la descripción del *teórico*; no *damos por sentado* que las características estructurales del organismo en las cuales el teórico basa su descripción incluyan elementos que son descripciones en sí mismos. [Las características del deshuesador de cerezas que nos llevan a describir una cereza (en lugar de un melocotón o un aceituna) no son descripciones de cerezas en sí mismas.] Desde esta perspectiva vemos que Putnam inventó el Planeta Tierra Gemelo y la Tierra para que ambos fueran miembros del conjunto de mundos posibles que *es* el mundo nocional que comparto con mi *Doppelgänger*. XYZ mitiga la sed, disuelve la pasta de empapelar y produce arcos iris tal como lo hace el H_2O; su diferencia con el H_2O está por debajo de todos los umbrales de discriminación tanto míos como los de mi *Doppelgänger*, siempre que, presumiblemente, ninguno de nosotros esté en contacto con, o consulte a un químico o microfísico taimado.

Dado un mundo nocional para un sujeto, podemos hablar *acerca* de qué son las creencias del sujeto, en un sentido peculiar pero conocido del "acerca de". Goodman (1961) discute oraciones de Dickens que son "acerca de Pickwick", un rasgo semántico de estas oraciones que no es auténticamente relacional al no haber ningún señor Pickwick de quien ellas hablan en el sentido relacional fuerte. En un espíritu similar Brentano discute el estado "tipo relación" de los fenómenos mentales cuyos objetos intencionales son inexistentes (véase Aquila, 1977). Una suposición habilitante de la psicología de la actitud nocional es que el teórico puede utilizar la acerquidad pickwickiana y sus semejantes como las propiedades semánticas que hacen falta para los fundamentos de cualquier teoría de la representación mental.

No es que la estrategia no esté probada. Aunque la psicología de la actitud nocional se ha tramado aquí como una respuesta a los problemas filosóficos encontrados en la psicología de la actitud proposicional y de la actitud

---

demás que abarcan esas especificaciones: decir que mi casa de ensueño está construida con cedro no es decir que mi especificación está hecha de cedro, sino que cualquier casa construida de acuerdo con mi especificación lo estaría. Decir que proyecto construir*la* el año próximo es decir que proyecto construir una casa de acuerdo con mis especificaciones el año próximo.

[14] Características especiales de la ficción (literaria) llevaron a Lewis a hacerle sustanciales modificaciones ingeniosas a esta idea, para explicar el papel de las presunciones de fondo, los conocimientos del narrador y demás en la interpretación normal de la ficción. Por ejemplo, damos por sentado que el mapa de Londres de Holmes es el del Londres victoriano, excepto allí donde está desbordado por los inventos de Conan Doyle; los textos ni afirman ni implican estrictamente que Holmes no tenía una tercera fosa nasal, pero los mundos posibles en los que éste sería el caso están excluidos.

oracional, se puede discernir fácilmente que es la metodología tácita y la ideología de una rama mayor de la Inteligencia Artificial. Consideremos, por ejemplo, el ahora famoso sistema SHRDLU de Winograd (1972). SHRDLU es un "robot" que "vive en" un mundo que consiste en una mesa sobre la que hay cubos de distintos colores y formas. El los percibe y los mueve en respuesta a órdenes mecanografiadas (en inglés) y puede contestar preguntas (en inglés) acerca de sus actividades y el estado de su mundo. Las citas que sustentan de más atrás son cruciales, puesto que SHRDLU no es en realidad un robot y no hay ninguna mesa con cubos para que SHRDLU los manipule. Ese mundo y las acciones de SHRDLU en él están meramente simuladas en el programa del ordenador del cual SHRDLU, el *simulacro* de robot es parte. Fodor (1980) formula la proposición que queremos, hasta anticipando nuestra terminología:

> En efecto, la máquina vive en un mundo completamente nocional; todas sus creencias son falsas. Por supuesto, a la máquina no le importa que sus creencias sean falsas puesto que la falsedad es una propiedad semántica y, *qua* ordenador el mecanismo satisface las condiciones formales; viceversa, sólo tiene acceso a las propiedades formales (no semánticas) de la representación que manipula. En efecto, el mecanismo está precisamente en la situación que Descartes teme: es sólo un ordenador que sueña que es un robot.

Para algunos críticos el hecho de que el SHRDLU no perciba en realidad las cosas del mundo, que no las toque, ni de otro modo entre en relaciones causales con ellas, es suficiente para demostrar que, sea lo que sea, lo que SHRDLU tiene, por cierto carece en absoluto de *creencias*. ¿Qué creencias podría tener SHRDLU? ¿Cuál podría ser su contenido? ¿Acerca de qué podrían ser? SHRDLU es un sistema puramente formal del todo desligado del mundo por lazos de percepción, acción, o por cierto de interés. ¡La idea de que esos estados y procesos meramente formales, meramente sintácticos, totalmente carentes de propiedades semánticas nos puedan proporcionar un modelo de creencia, es indignante! (SHRDLU provoca la fanfarronería de la gente).

La respuesta amable de la que se dispone en principio es la siguiente. Por cierto como lo proclaman los críticos, un auténtico creyente debe estar rica e íntimamente ligado por la percepción y la acción a las cosas del mundo, los objetos de sus creencias, pero proporcionarles esos lazos al SHRDLU mediante los ojos de una cámara de TV verdadera, un brazo robótico real y una mesa verdadera con cubos en la cual vivir, habría sido caro, habría desperdiciado tiempo y habría tenido *poco interés psicológico*. Cubierto por un transductor, abrigo efector del Hardware robótico SHRDLU tendría un mundo nocional de cubos en la tabla de una mesa, lo que equivale a decir que arrojado a ese ambiente *real* ese SHRDLU se las arreglaría muy bien; un mundo de cubos es un buen hielo para SHRDLU. Despojado del abrigo robótico, SHRDLU tiene un mundo nocional ampliamente menos específico; muchos más mundos posibles son imperceptibles para él, pero sin embargo la estructura funcional del núcleo es el habitáculo de los interesantes problemas psicológicos y sus soluciones propuestas, de manera que elegir el mundo de

los cubos como mundo nocional *admisible* (está en el conjunto de los modelos de Tarski para el sistema del núcleo) es una manera inocente de cubrir el sistema con alguna verosimilitud.

En el caso real del SHRDLU, esta defensa sería optimista: SHRDLU no es así de bueno. *No* sería trivial, ni siquiera una forma cara pero directa de ingeniería, cubrir a SHRDLU con robótica, y las razones de por qué no lo sería son de interés psicológico. Al mantener el mundo de SHRDLU meramente nocional, Winograd se disculpó claramente de dar soluciones a una gran cantidad de problemas difíciles, profundos e importantes de la psicología. Dista mucho de ser claro que cualquier mejora en el SHRDLU concebida en el mismo espíritu dentro del mismo programa de investigación pudiera aprovecharse con justicia de esta línea de defensa. Pero creo que es indiscutible que es la supuesta línea de defensa ideal de esa investigación. Ante la aseveración de Husserl de que separar el mundo real nos deja con la esencia de lo mental, Winograd y Al pueden agregar: sí, y además separar ahora tiempo y dinero.

Los temas husserlianos en este programa de investigación de *Al* son inequívocos, pero es importante que recordemos igualmente las diferencias. Para el autofenomenólogo, la relativa inaccesibilidad de los referentes verdaderos de las creencias de alguien —y por tanto, como arguye Putnam, la relativa inaccesibilidad de las actitudes *proposicionales* de alguien – se presenta como un punto acerca de los límites del acceso introspectivo privilegiado, un resultado muy cartesiano: *yo* no puedo discriminar *con seguridad*; no tengo *autoridad* acerca de qué proposición me estoy ocupando, en qué objeto real estoy pensando. Pero las "introspecciones" de SHRDLU no tienen ningún papel privilegiado en la heterofenomenología de Winograd: el mundo nocional de SHRDLU se fija desde el exterior apelando a hechos objetivos y públicamente accesibles sobre las capacidades y disposiciones del sistema y por tanto su destino en distintos ambientes imaginados. La contrafigura de la afirmación cartesiana es que hasta la totalidad de estos hechos *públicos* subdetermina las actitudes proposicionales. Aun cuando los ambientes a los que se apela sean imaginarios, la apelación coloca a la heterofenomenología directamente del lado naturalista de la división Jamesiana.

La elaboración de ambientes ideales imaginarios a los fines de comparar sistemas internamente diferentes es una estrategia de cierta actualidad, por ejemplo en la ingeniería. Podemos comparar la potencia de distintos motores de automóvil imaginándolos en concursos de tiro contra cierto caballo ficticio, o podemos comparar la eficiencia de sus combustibles viendo cuán lejos llegarán con un coche en determinado ambiente simulado. El uso de un ambiente ideal nos permite describir semejanzas o competencias *funcionales* independientemente de los detalles de implementación o desempeño. Utilizar la estrategia en psicología para elaborar mundos nocionales es simplemente un caso especialmente complejo. Nos permite describir semejanzas parciales en las "competencias" psicológicas de sujetos diferentes —por ejemplo, sus poderes representativos – de maneras neutrales en lo que respecta a su implementación: por ejemplo, sus *medios* representativos.

La analogía con la ficción vuelve a ser útil para hacerse entender en este

punto. ¿Cuál es exactamente la semejanza entre el *Romeo y Julieta* de Shakespeare y *Amor sin barreras* de Bernstein? La segunda estaba "basada en" la primera, como sabemos, pero ¿qué tienen en común en realidad? ¿Se refieren a la misma gente? No, puesto que ambas son ficción. ¿Contienen las mismas o similares representaciones? ¿Qué significaría esto? ¿Las mismas palabras u oraciones o descripciones? Los libretos de ambas están escritos en inglés, pero esto es en verdad poco importante, puesto que la semejanza que perseguimos sobrevive en la traducción a otros idiomas, y —más dramáticamente— es evidente en la película "West Side Story" y en la ópera de Gounod. La semejanza es independiente de cualquier medio determinado de representación —libretos, *sketches*, descripciones, actores en escenarios o delante de las cámaras— y concierne a *lo que se representa*. No es una semejanza sintáctica de ninguna clase. Puesto que tales semejanzas son tan evidentes en la ficción como en el informe real, debemos comprender que "lo que se representa" nos lleva a elementos de un mundo nocional, no necesariamente el mundo real. Podemos comparar diferentes mundos reales o de ficción con relación a asuntos grandes y pequeños, tal como podemos comparar distintas partes del mundo real. Podemos comparar un mundo nocional con el mundo real (el mundo nocional del miope Mister Magoo se parece al mundo real sólo de manera intermitente y parcial, pero lo bastante, milagrosamente, como para salvarlo del desastre).

¿Entonces, cuándo diremos que dos personas diferentes comparten una actitud nocional o conjunto de actitudes nocionales? Cuando sus mundos nocionales tienen un punto o área de semejanza. Los mundos nocionales son centrados en el agente o egocéntricos (Perri, 1977; Lewis, 1979); al comparar mundos nocionales en busca de la semejanza psicológica será útil, por lo tanto, "superponer" los centros —de manera que los orígenes, la intersección de los ejes, coincida— antes de probar la semejanza. De este modo, surgirán las semejanzas psicológicas entre dos paranoides, mientras que la diferencia psicológica entre el masoquista y su compañero sádico resalta a pesar de la gran similitud en los *dramatis personae* de sus mundos nocionales cuando se los estudia descentrados.[15]

La perspectiva de un método riguroso de comparación del mundo nocional —un procedimiento de decisión para encontrar y evaluar los puntos de coincidencia, por ejemplo— es confusa. Pero siempre hemos sabido eso, puesto que las perspectivas de establecer condiciones para la identidad de la actitud proposicional son igualmente vagas. Creo que la sal es cloruro de sodio, pero mis conocimientos de química son deficientes; el químico también cree que la sal es cloruro de sodio, pero no va a haber ninguna forma eficaz de captar el núcleo central de nuestras creencias (Dennett, 1969). La comparación de creencias, vistas como actitudes nocionales o proposicionales, no se va a convertir en rutina por medio de un golpe teórico. La *mejor precisión* que se podría equivocadamente haber esperado lograr aislando y traducien-

[15] Los puntos que rodean a "Yo" y la indicatividad son mucho más complicados que lo que revela este reconocimiento apresurado. Véanse no sólo Perry y Lewis sino también Castañeda (1966, 1967, 1968). Para reflexiones ilustrativas sobre un tema similar véase Hofstadter, 1979, págs. 373-76.

do el "lenguaje del pensamiento" —si existe— no mejoraría la comparación de las *creencias*, tales como la mía y la del químico acerca de la sal, sino sólo la comparación de cierto tipo innovador de oración: las oraciones en la cabeza. Pero las oraciones ya son fácilmente comparables. El químico de habla inglesa y yo usamos exactamente las mismas palabras para expresar nuestra creencia acerca de la sal, y si por casualidad nuestros cerebros hacen lo mismo, seguiremos teniendo el problema de la comparación para nuestras creencias.

Un lenguaje del pensamiento no daría ningún poder mayor en el caso enojoso de creencias irracionales y especialmente contradictorias, y por la misma razón. Supongamos que se divulga la hipótesis de que Bill tiene un par especial de creencias contradictorias: cree tanto que Tom es digno de confianza como que Tom es indigno de confianza. En cualquier idioma digno de su nombre nada es más rutinario que determinar cuando una oración contradice otra, de manera que si conocemos el lenguaje del pensamiento de Bill, buscamos el par pertinente de oraciones en su cerebro ¡y las encontramos! ¿Qué demostraría esto? La pregunta seguiría siendo: ¿en cuál cree si es que cree en alguna? Haciendo una investigación ulterior podríamos descubrir que una de estas oraciones era rudimentaria y no funcional, que nunca había sido borrada de la pizarra del cerebro pero que tampoco se la había consultado nunca. O podríamos descubrir que una oración (en mentalés para "Tom es indigno de confianza") fue consultada y (actuada) de manera intermitente, buena prueba de que Bill cree que Tom no es digno de confianza, pero que lo sigue olvidando. Se olvida, y luego su natural bonhomía se hace cargo, y al creer que la gente en general es digna de confianza, se conduce como si creyera que Tom también lo es. O tal vez encontremos un comportamiento verdaderamente conflictivo en Bill; sigue y sigue conversando sobre la integridad de Tom, pero observamos que nunca le da la espalda. Se pueden multiplicar los casos, llenando espacios y extendiendo los extremos, pero en ninguno de ellos la presencia o ausencia de contradicción explícita en el lenguaje mentalés desempeña realmente un papel secundario más que periférico en nuestra decisión de caracterizar a Bill como vacilante, olvidadizo, indeciso o verdaderamente irracional. La conducta de Bill da para más, pero la *conducta* tampoco resolverá el caso (véase el capítulo 4 y sus reflexiones).

La gente en verdad se confunde y aún peor; a veces se vuelve loca. Decir que alguien es irracional es decir (en parte) que en cierto sentido está mal equipado para tratar con el mundo en el que habita; no se adapta bien a su espacio. En los casos graves podemos ser incapaces de crearle algún mundo nocional; ningún mundo posible sería el lugar en el que encajara bien. Se podría dejar el tema ahí, o intentar ser más descriptivo de la confusión de esa persona.[16] Se podría componer una descripción reconocidamente inconsistente, citando el capítulo y el verso de las propensiones y la constitución in-

[16] "Un hombre puede pensar que cree *p* al mismo tiempo que su conducta sólo puede explicarse por medio de la hipótesis de que cree no-*p*, dado que es sabido que quiere *z*. Tal vez la confusión que hay en su mente no puede ser transmitida por ninguna explicación simple (o compleja-D.C.D.) de lo que él cree: quizá sólo una reproducción de la complejidañd y la confusión sería exacta" (Hampshire, 1975, pág. 123).

terna de la conducta del sujeto en apoyo de las distintas partes de la descripción. Una descripción tan inconsistente no podría ser la de *un* mundo nocional, puesto que los mundos nocionales, como conjuntos de mundos *posibles*, no pueden tener propiedades contradictorias, pero no hay nada que garantice que un sujeto tiene un único mundo nocional coherente. Su mundo nocional puede estar partido en mundos fragmentarios, superpuestos, competitivos.[17] Cuando el teórico, el heterofenomenólogo o el psicólogo del mundo nocional, siguen la alternativa de ofrecer una descripción reconocidamente inconsistente de un mundo nocional, esto no cuenta como una caracterización establecida, positiva de un mundo nocional, sino como una rendición frente a la confusión, abandonando la tentativa de la interpretación (completa). Es igual que caer en el lenguaje directo cuando se quieren transmitir los comentarios de alguien. "Y lo que él *dijo* fue: 'Nada *nadea*'".

La heterofenomenología del mundo nocional, no resuelve, entonces, las disputas e indefiniciones y ni siquiera aguza los límites de lo que la gente común piensa acerca de las creencias; hereda los problemas y simplemente los reconstruye en un formato levemente nuevo. Uno podría muy bien preguntarse qué recursos tiene que lo recomienden. La perspectiva de construir el nocional de un ser? Trabajar en la otra dirección empezando por la descripes sumamente remota, de manera de ¿qué valor puede concebir el mundo nocional de un ser? Trabajar en la otra dirección empezando por la descripción de un mundo nocional y preguntando después cómo diseñar un "ser" que tenga ese mundo nocional. Parte del atractivo de *AI* es que proporciona un modo de empezar por lo que son esencialmente categorías y diferencias fenomenológicas —características de los mundos nocionales— y retroceder hasta hipótesis sobre cómo implementar esas aptitudes. Se comienza por los *poderes* representativos y se trabaja en dirección a los *medios*. Los filósofos también han jugueteado con esta estrategia.

La literatura filosófica reciente acerca de la diferencia entre las creencias y otras actitudes *de re* y *de dicto* está llena de sugerencias incompletas para distintos tipos de mecanismos mentales que podrían jugar un papel crucial en señalar esa diferencia: los *nombres vívidos* de Kaplan (1968), los *modos de presentación* de Schiffer (1978) y varios otros autores, y los *aspectos* de Searle (1979), por nombrar algunos. Se supone que estos son puramente definibles en los términos de la psicología restringida,[18] de manera que la psicología de la actitud nocional, debería, en principio, ser capaz de captarlas. Cuando nos volvamos hacia esa literatura en la próxima sección, exploraremos las perspectivas de esos mecanismos, pero primero hay más fundamentos para el escepticismo que sacar a la luz acerca de los mundos nocionales.

El tema de un mundo nocional, un mundo *constituido* por la mente o la experiencia de un sujeto, ha sido un *leitmotiv* recurrente en la filosofía por lo

[17] Véase la discusión de la Fenomenología y la "Feenomanology" en "Two Approaches to Mentales Images", en *Brainstorms*. Véanse también los comentarios de Lewis (1978) acerca de cómo tratar la inconsistencia en una obra de ficción.

[18] Kaplan es explícito (1968):"El rasgo crucial de esta noción (los nombres vívidos de Ralph) es que depende sólo del presente estado mental de Ralph, y pasa por alto todos los vínculos ya sea por semejanza o génesis con el mundo real... Está destinado a llegar a los aspectos puramente internos de la individualización" (pág. 201).

menos desde Descartes. De distintas maneras ha perseguido el idealismo, el fenomenalismo, el verificacionismo y la teoría coherente de la verdad, y, a pesar de las palizas que típicamente recibe, sigue resucitando en versiones nuevas y mejoradas: en *Ways of World-making* de Goodman (1978) y en la reciente reevaluación del realismo en Putnam (1978), por ejemplo. La ubicuidad del tema no es prueba de su solidez en ningún aspecto; puede no ser más que un error siempre tentador. En su aspecto actual corre temerariamente hacia una intuición igualmente compulsiva acerca de la referencia. Si las actiudes nocionales han de jugar el papel intermediario que les fue asignado, si han de ser la contrafigura para la psicología del concepto del carácter de una expresión lingüística de Kaplan, debe inferirse que cuando a un sujeto psicológico o ser, con su mundo nocional determinado por su constitución interna se lo coloca en distintos contextos, en diferentes ambientes verdaderos, esto debería determinar diferentes actitudes proposicionales para el sujeto.

actitud nocional + ambiente → actitud proposicional

Eso significa que si yo y mi *Doppelgänger* fuéramos cambiados, instantáneamente, (o en todo caso sin permitir que ocurra ningún cambio del estado interno durante la transición, el intercambio podría llevar tanto tiempo como si yo y mi *Doppelgänger* estuviéramos en coma durante todo ese tiempo) yo despertaría con actitudes proposicionales *acerca de las cosas en el Planeta Tierra Gemelo*, y mi *Doppelgänger* tendría *actitudes* proposicionales *acerca de las cosas de la Tierra.*[19] Pero eso es muy poco intuitivo para mucha gente, descubro, pero no para todos. Por ejemplo, yo tengo muchas creencias y otras actitudes acerca de mi esposa, una persona de la Tierra. Cuando mi *Doppelgänger* se despierta por primera vez en la Tierra después del cambio, y piensa "Quisiera saber si Susan ya ha hecho el café" *con seguridad* no está pensando pensamientos acerca de *mi esposa.* ¡Nunca la conoció ni oyó hablar de ella! Seguramente sus pensamientos se refieren a *su* Susan, años luz atrás, aunque con seguridad él no tiene la menor idea de la distancia. El

[19] Sin embargo mi *Doppelgänger* no tendría pensamientos *acerca de mí* cuando pensaba "Tengo sueño", y así sucesivamente. La referencia al pronombre de la primera persona no se ve afectada por el cambio de palabras, por supuesto (véase Putnam, 1975a; Perry, 1977, 1979; Lewis, 1979). Pero hay que tener cuidado en no inflar este punto hasta convertirlo en una doctrina metafísica acerca de la identidad personal. Consideremos esta variación sobre un tema de ciencia ficción conocido en filosofía. Su nave espacial se estrella en Marte y usted quiere volver a la Tierra. Afortunadamente, hay disponible un telecargador. Usted sube a la cabina en Marte y él le hace un análisis microfísico completo que exige disolverlo a usted en sus átomos componentes, por supuesto. Irradia la información a la Tierra, donde el receptor, provisto de montones de átomos, del mismo modo en que una fotocopiadora está provista con papel blanco nuevo, crea un duplicado exacto de usted que baja y "continúa" su vida en la Tierra con su familia y amigos. ¿El telecargador "asesina para disecar"?, ¿o lo ha llevado a usted a su casa? Cuando el terráqueo – usted recién llegado dice "Tuve un accidente muy feo en Marte", ¿es cierto lo que dice? Supongamos que el telecargador pueda obtener su información acerca de usted sin disolverlo, de manera que usted continúe una vida solitaria en Marte. En sus marcas, listos, ya... (algunos de los que han adoptado este juego de fiesta filosófica: Hofstadter y Dennett, 1981; Nozick, 1981; Parfit, 1984; Nagel, 1986).

hecho de que él no se entere nunca de la diferencia, como tampoco ninguna otra persona, salvo el Demonio Malvado que efectuó el cambio, carece de importancia; lo que nadie podría verificar sería no obstante, cierto: sus pensamientos no se refieren a mi esposa, al menos no hasta que él haya tenido algún contacto causal con ella.

Esto es, en esencia, la teoría causal de la referencia (véase, por ej., Kripke, 1972; Evans, 1973; Donellan, 1966, 1970, 1974) y el experimento del pensamiento la aísla muy bien. Pero las intuiciones provocadas en circunstancias tan alocadas de ciencia ficción son una prueba pobre. Consideremos el mismo punto como podría suceder en una secuela de hechos perfectamente posibles aquí en la Tierra. En Costa Mesa, California, hay, o por lo menos había, un establecimiento llamado la Pizzería de Shakey, un lugar extravagante que anunciaba una pianola desafinada con teclas fluorescentes, y con distintos letreros "graciosos" pintados a mano en las paredes: "Shakey ha hecho un trato con el banco: nosotros no aceptamos cheques y el banco no hace pizza", etcétera. Por extraño que parezca, muy extraño en realidad, en Wetwood Village, California, a unos ochenta kilómetros de distancia, había otra Pizzería de Shakey, que era misteriosamente parecida: construida de acuerdo con los mismos planos, la misma pianola desafinada, los mismos letreros, la misma playa de estacionamiento, el mismo menú, las mismas mesas y bancos. Cuando noté eso por primera vez se me ocurrió la broma obvia, pero, triste es decirlo, jamás la llevé a cabo. Se la habría podido hacer con facilidad, sin embargo, así que permítanme que les cuente el cuento como si realmente hubiera sucedido.

La balada de la Pizzería de Shakey

Una vez Fulano, Zutano y Mengano fueron a lo de Shakey en Costa Mesa a comer pizza con cerveza, y Zutano y Mengano le hicieron una broma a Fulano, que era nuevo en la zona. Después de haber pedido su comida y empezado a comer, Fulano fue al baño de hombres, y en ese momento Zutano deslizó un soporífero en la cerveza de Fulano. Fulano volvió a la mesa, vació su jarro, y muy pronto se quedó completamente dormido junto a la mesa. Zutano juntó las pizzas que no habían sido comidas. Mengano descolgó el sombrero de Fulano del colgador que tenía detrás de la cabeza y luego arrastró a Fulano afuera y al coche, y condujo a toda velocidad hasta Westwood Village, donde se volvieron a instalar, con una nueva jarra de cerveza y algunos jarros, en la mesa idéntica. Luego se despertó Fulano. "Debo de haberme quedado dormido", comentó, y la noche siguió ruidosamente, como antes. La conversación giró al tema de los letreros y otros aspectos de la decoración y luego al de las inscripciones en las paredes; para deleite de Zutano y Mengano, Fulano señaló en dirección al baño de hombres y confesó que aunque no pertenece en realidad a esa clase de hombres, esa noche se había sentido inspirado como para inscribir sus iniciales en la puerta del retrete que estaba más hacia la izquierda en ese baño de hombres. Zutano y Mengano dudaron de su palabra, y entonces Fulano propuso una apuesta. Anunció

que estaba dispuesto a apostar que sus iniciales estaban inscritas en esa puerta. Zutano aceptó la apuesta, con Mengano de árbitro, y sacaron lápiz y papel, en el que escribirían la expresión explícita del punto en discusión. En este punto, el *suspense* era muy grande, porque el que Zutano ganara o no la apuesta dependía de las palabras exactas. Si Fulano escribía: "Apuesto cinco dólares a que mis iniciales aparecen en la puerta del retrete que está más a la izquierda del baño de hombres de la Shakey de Costa Mesa", Fulano ganaría la apuesta. Pero si escribía: "Apuesto cinco dólares a que mis iniciales aparecen en el retrete que está más a la izquierda del baño de hombres de la pizzería en la que estamos sentados ahora" —u otras palabras en *ese* sentido— ganaría Zutano. Una tercera posibilidad era que Fulano armara una oración que *no expresara una proposición* porque contuviera un nombre o una descripción vaga: "la Shakey de Costa Mesa en la que estamos sentados ahora" o "el baño de hombres del lugar en el que compré y consumí totalmente una pizza de anchoas la noche del 11 de febrero de 1968". En tal caso Mengano se vería obligado a declarar que la apuesta estaba mal y devolvería el dinero. (Si Mengano es un Russeliano estricto en lo que se refiere a las descripciones estrictas, puede declarar falsa la oración de Fulano en estos casos, y premiar a Zutano con el dinero).

Pero Fulano quedó a merced de ellos, comprometiéndose con la puerta de Westwood Village por escrito (por supuesto no según *esa* descripción), y perdió la apuesta. Le explicaron la broma y, aunque admitió haber sido burlado, reconoció que se había comprometido con una proposición falsa y perdió limpiamente la apuesta. Pero ¿qué puerta había tenido en la mente? Pues bien, él podía insistir en que en algunos aspectos, había estado pensando en la puerta de Costa Mesa. Había recordado vívidamente el episodio de su cortaplumas hundiéndose en esa puerta. Pero también se le había "presentado" vívidamente la puerta como si estuviera a poca distancia y había anticipado ansiosamente en su imaginación, el triunfo que sobrevendría cuando ellos tres entraran en el baño adyacente para dirimir la apuesta. De manera que había también mucho que decir en favor del hecho de que hubiera pensado en la puerta de Westwood Village. ¡Qué enigma! En verdad éste fue un trabajo para filósofos sobrios con un vocabulario técnico a su disposición.

Los filósofos dicen que hay una diferencia entre la creencia *de re* y la creencia *de dicto*. Todos conocen esta diferencia en el fondo de su corazón, pero como en muchas diferencias filosóficas importantes, es muy difícil caracterizar de manera precisa e indiscutible. Estamos estudiándolo. Mientras tanto, señalamos la diferencia que tiende a perderse en la ambigüedad de la conversación casual, usando siempre el estilo extraño pero al menos discutible de la atribución, cuando hablamos de las creencias *de re*, reservando el estilo "que" para las atribuciones de creencia *de dicto*.[20] Así

1) Bill cree *que* el capitán del equipo soviético de hockey sobre hielo que es hombre
pero no es el caso de que

[20] Los expertos de la literatura notarán que esta oración reproduce delicadamente una equivocación conocida que indica el *género*: ¿modifica el *de re* el "habla" o las "creencias; modifica el *de dicto* las *"atribuciones"* o la "creencia".

2) Bill crea *del* capitán del equipo soviético de hockey sobre hielo que es hombre
puesto que Bill no conoce nada a ese ruso fornido, quienquiera que sea. Por contraste,

3) Bill cree *de* su propio padre que es hombre.

Con seguridad todos conocemos *esta* diferencia, la distinción ostentada por el ejemplo, de manera que ahora pasamos a aplicarlo en el caso de la Pizzería de Shakey. En virtud del rico intercambio causal de Fulano con la Shakey de Costa Mesa y lo que hay en ella, Fulano tiene derecho a creencias *de re* que lo relacionan con esas cosas. Cuando despierta en Westwood Village a medida que su mirada recorre el lugar rápidamente recoge las relaciones causales obligatorias con *muchos* de los objetos de Westwood Village también, incluyendo a la Pizzería Shakey de Westwood Village misma. De este modo podemos catalogar algunas de las creencias *de re* verdaderas y falsas que Fulano tenía poco después de despertar.

Fulano cree *de* la Shakey de Costa Mesa:

Verdadero	*Falso*
que esta noche compró una pizza allí	que está allí ahora
que se quedó dormido allí	que despertó allí
que colgó su sombrero en un perchero de allí	que su sombrero está en el perchero allí

Fulano cree *de* la Shakey de Westwood Village:

Verdadero	*Falso*
que está allí ahora	que compró una pizza allí esta noche
que despertó allí	que se quedó dormido allí
que su sombrero está en un perchero de allí	que él cuelga su sombrero en un perchero de allí

Allí donde en el curso normal de las cosas una persona tendría una lista única de creencias *de re*, Fulano, debido a la dislocación ingeniosa que hemos producido, tiene una lista dual de creencias *de re*; cada creencia *de re* verdadera tiene una falsa gemela acerca de un objeto diferente. Por supuesto que Fulano desconoce completamente esta duplicación de sus creencias; todavía hay *algo unitario* acerca de su *estado psicológico*. (Podríamos decir: hay unidad en su mundo nocional. Donde hay dualidad es en el mundo real. Cada una de sus actitudes nocionales únicas engendra un par de actitudes proposicionales, dadas sus peculiares circunstancias.)

Pero los problemas surgen cuando intentamos continuar la lista de las creencias *de re* de Fulano. El presumiblemente cree *de* su sombrero tanto que está en un perchero en Costa Mesa como que está en un perchero justo detrás de su cabeza. Habiendo notado el perchero de Costa Mesa, por más casualmente que haya sido, se puede decir que Fulano también cree *del* perchero de Costa Mesa que su sombrero está en él (o, juntando ambas cosas:

cree de su sombrero y de ese perchero que el primero está en el segundo). ¿Pero es posible que crea del perchero que tiene detrás de su cabeza, con el cual su única interacción causal hasta la fecha ha sido una atracción gravitacional mutua infinitesimalmente débil, que su sombrero está en él? El teórico causal debe negarlo. Se pensaría que el estado psicológico de Fulano *vis-à-vis* su sombrero y su colocación era muy sencillo (y así le parece a Fulano), pero en realidad es maravillosamente *complejo* cuando se lo somete al análisis filosófico. Fulano cree *que* su sombrero está en un perchero que tiene detrás de la cabeza (y ésa es una creencia verdadera); también cree realmente *que* el perchero q tiene detrás de la cabeza y en el que se apoya su sombrero está hecho de madera. El no cree *de* ese perchero, sin embargo, que esté hecho de madera ni que lo tenga detrás de la cabeza. Más aun, Fulano cree realmente *que* hay una puerta en el retrete más a la izquierda del baño de hombres adyacente y cree falsamente *que* esa puerta tiene sus iniciales grabadas en ella. De manera que cree que la puerta del retrete más a la izquierda tiene sus iniciales, pero no cree de esa puerta que tiene sus iniciales.

Algunos filósofos disentirían. Algunos (por ej., Kaplan, 1968) dirían que la afinidad de Fulano con el perchero de Costa Mesa era demasiado *casual* (si bien causal) como para calificar a Fulano por sus creencias *de re* acerca de él. Avanzando en la otra dirección algunos (por ej., Kaplan, 1968) confiarían en debilitar el requisito causal (y reemplazarlo con otra cosa, aún por determinar) de modo que Fulano pudiera tener creencias *de re* acerca del perchero que no vio y la puerta no escrita. Y algunos se plantarían en sus trece y alegarían que las diferencias reconocidamente exóticas extraídas en el párrafo anterior no eran nada más que las implicaciones tolerables de una buena teoría colocada *in extremis* por condiciones muy desusadas.

El sentido de proseguir en estos desacuerdos, de zanjar la disputa filosófica, podría muy bien perdérsele a un psicólogo. Es tentador sostener que los problemas *filosóficos* encontrados aquí, si bien son problemas serios y reales cuya solución vale la pena buscar, no son para nada problemas para la psicología. Porque obsérvese que las distintas escuelas de pensamiento sobre las creencias *de re* de Fulano no difieren en las predicciones que harían acerca de la conducta de Fulano en distintas circunstancias. *Cuáles* oraciones pueden resultar seductoras para apostar por ellas, por ejemplo, no depende de qué creencias *de re* tiene él *verdaderamente*. Ninguna escuela puede proclamar una superioridad predictiva basada en su catálogo más exacto de las creencias de Fulano. Aquellas que sostienen que él no tiene creencias *de re* acerca de la puerta que no vio, describirán *retrospectivamente* aquellos casos en los que Fulano hace una apuesta perdedora como casos en que él de todos modos afirma algo que no tiene intención de afirmar, mientras que aquellos que tienen la convicción opuesta lo tendrán en cuenta en aquellas ocasiones como habiendo expresado (de todos modos) exactamente lo que creía. En el caso imaginado, tal vez no en otros casos más normales, la presencia o ausencia de una creencia *de re* determinada no juega ningún papel predictivo, por tanto no explicativo. Pero si en el caso imaginado no juega ningún papel, ¿no tendríamos que abandonar el concepto en favor de algún otro que pueda caracterizar las variables cruciales tanto en los casos normales como en los anormales?

El fracaso aparente de las distinciones filosóficas para enredarse con cualquier diferencia filosófica útil puede deberse, sin embargo, a que estemos buscando en el lugar equivocado, centrándonos demasiado ajustadamente en una imperceptibilidad *local* inventada y perdiendo de este modo la importante diferencia psicológica que surge de algún modo, en un contexto más amplio. La familia de los casos extravagantes tramados por quienes toman parte en la literatura con la inclusión de bromas complicadas, juegos con espejos, personas disfrazadas de gorilas, gemelos idénticos y el resto de los trucos teatrales destinados a producir casos de *falsa identidad*, logran producir sólo momentáneamente o en el mejor de los casos efectos inestables de la clase deseada. No es fácil sostener la clase de ilusión necesaria para cimentar los veredictos anónimos u otros enigmas. Extraer veredictos basados en las anomalías de poca duración en el estado psicológico de una persona dan una imagen seriamente distorsionada del modo en que la gente está relacionada con las cosas del mundo; nuestra capacidad para seguir el rastro de las cosas a través del tiempo no está bien descrita por ninguna teoría que atomice los procesos psicológicos en momentos sucesivos con ciertas características.[21] Creo que todo esto es muy plausible, pero, ¿qué conclusión habría que sacar? Tal vez ésta: la semántica formal nos exige determinar el objeto a evaluar en un momento especial y en un contexto especial para el valor de la verdad o la referencia, y mientras que la conducta lingüística abierta le proporciona al teórico objetos candidatos —elocuciones— para ese papel, efectuando movidas internas y postulando objetos o estados "mentales" análogos para esa determinación, debe ejercer violencia sobre la situación psicológica. Quienquiera que importe las categorías necesarias para una teoría semántica formal y las obligue a servir en una teoría psicológica, está destinado a crear un monstruo. Tal conclusión es, como diría James, "la vaguedad personificada". En especial, no está claro todavía si podría ser una conclusión tan fuerte como para amenazar *todas* las versiones de la teoría de la "representación mental", todas las teorías que suponen que en la cabeza hay objetos sintácticos para los que se puede dar una interpretación semántica de principios.

Me resulta muy difícil expresar de manera más precisa esta preocupación, pero por el momento puedo lograr hacerla más vívida con la ayuda de una analogía. Una de las escenas cómicas más inspiradas que se transmiten regularmente en el espectáculo televisivo *Laugh-In* era el "teatro robot" en el que Arte Johnson y Judy Carne representaban a un par de robots recién casados. Aparecían en alguna circunstancia mundana, preparando el desayuno o "el maridito está de vuelta de la oficina" y se movían en un simulacro levemente convulsivo de la acción humana. Pero las cosas nunca resultaban completamente bien: Arte tendía la mano para abrir una puerta, no alcanzaba el picaporte por poco, giraba la muñeca, revoleaba el brazo y se estrellaba de cabeza contra la puerta aún cerrada; Judy le servía café a Arte

[21] En conferencias dictadas en Oxford en 1979, Evans desarrolló el tema del *proceso* de mantenerse al tanto de las cosas del mundo (véase Evans, 1982). Se hace eco de un tema central en el renunciamiento apostático de los experimentos bidimensionales taquistoscópicos de Neisser (1976) en la psicología de la percepción, a favor de un enfoque "ecológico" gibsoniano de la percepción.

pero el café erraba la taza —no tenía importancia puesto que Arte no se daba cuenta— y "vaciaba" la taza, se volvía amorosamente hacia Judy y le decía "¡delicioso!". Etcétera. Se veía que el "problema" era que sus *mundos nocionales no "coincidían" completamente con el mundo real;* se tenía la impresión de que si se los hubiera movido un par de centímetros antes de ponerlos "en marcha" todo hubiera andado como una seda; entonces sus creencias hubieran tenido la oportunidad de ser *acerca* de las cosas del mundo que los rodeaba. Presumiblemente, su conducta se explicaba por el hecho de que cada uno de ellos contenía una representación interna del mundo, consultando con el cual gobernaban su conducta. Así es como funcionan los robots. Esta representación interna se actualizaba constantemente, por supuesto, *pero no continuamente*. Sus mecanismos perceptivos (y los registros internos de sus acciones) les suministraban una sucesión de instantáneas, por decirlo así, de la realidad, que provocaba correcciones en sus representaciones internas, pero no lo bastante rápidas o exactas como para sostener una conjunción adecuada de su mundo nocional, su mundo tal como estaba representado y el mundo real.[22] De ahí la infelicidad de la conducta. El "chiste" es que *no somos así en absoluto*.

Pues bien, ¿lo somos o no lo somos? La esperanza de la ciencia cognitiva es que *seamos* así, sólo que mucho, mucho mejores. En apoyo de esta convicción la ciencia cognitiva puede señalar exactamente los casos anómalos considerados en la literatura acerca de las creencias *de re* y *de dicto*: en efecto, éstos no son más que experimentos que producen patología en el mecanismo y que son, en consecuencia, ricas fuentes de pistas acerca de los principios de diseño de ese mecanismo. Que haya que trabajar tanto para maquinar casos de patología real demuestra cuán buenos somos para poner al día nuestras representaciones internas. El proceso de estar al día con las cosas es prácticamente continuo, pero todavía tendrá una descripción clara en los términos del repaso rápido de un modelo interno. Además, como lo demuestran con frecuencia los casos patológicos, cuando se agrega un informe *verbal* al informe puramente *perceptivo* del sistema, las posibilidades de trastornos graves, de creación de objetos nocionales sin contrafiguras reales, y cosas por el estilo, aumentan dramáticamente. Las creencias adquiridas por medio del lenguaje crean problemas cuando se las debe hacer armonizar con creencias inducidas perceptivamente, pero éstos son problemas que se pueden solucionar dentro del campo de acción de la ciencia cognitiva Dennett, de próxima aparición c).

Esto sugiere que los problemas con los que nos topamos en el cuento de la Pizzería de Shakey provienen de la tentativa de aplicar un único conjunto de categorías a dos (o más) estilos muy diferentes de funcionamiento cognitivo. En uno de estos estilos, tenemos realmente representaciones internas de las cosas del mundo, cuyo contenido guía, de algún modo, nuestra conducta.

[22] El problema de preservar esta conjunción tiene como parte central el "problema del marco" de la Inteligencia Artificial que surge para planificar sistemas que deben razonar acerca de los efectos de las acciones consideradas. Véanse McCarthy y Hayes, 1969, y Dennett, 1978a, capítulo 7, y 1984c. Es o el problema más difícil que *AI* debe —y eventualmente puede— resolver o *de reductio ad absurdum* de la teoría de la representación mental.

En el otro estilo tenemos algo así como procedimientos para mantenernos informados acerca de las cosas del mundo, lo que nos permite minimizar nuestras *representaciones* de esas cosas, dejándonos consultar las cosas mismas, más que sus representantes, cuando necesitamos más información sobre ellas. Se pueden encontrar reflexiones acerca de este tema en la literatura de la filosofía (por ejemplo, Burge, 1977; Kaplan, 1968, 1978, 1980; Morton, 1975; Nelson, 1978), de la psicología (por ejemplo, Gibson, 1966; Neisser, 1976) y de la Inteligencia Artificial (por ejemplo, Pylyshyn, 1979), pero todavía nadie ha logrado desenredar las metas y presunciones de los distintos proyectos teóricos diferentes que convergen acerca del tema: la semántica del lenguaje natural, la semántica y metafísica de la lógica modal, la psicología cognitiva restringida de los individuos, la psicología amplia o naturalista de los individuos en los entornos y en los grupos sociales. Si nos armamos, tentativamente, con la idea de los mundos nocionales, que proporciona al menos una manera pintoresca, si bien no probadamente sólida de describir esos temas que pertenecen al campo de acción de la psicología restringida y de distinguirlos de los temas que requieren otra perspectiva, tal vez logremos alcanzar algún progreso por medio de la consideración de los orígenes de las diferencias problemáticas en el contexto de los problemas teóricos que los ocasionaron.

El *de re* y el *de dicto* desmantelados

> A menos que estemos preparados para decir que "la mente no puede llegar más allá del círculo de sus propias ideas, "debemos reconocer que algunas de las cosas del mundo pueden, en realidad, convertirse en objetos de nuestras actitudes intencionales. Uno de los datos acerca de Oliver B. Garrett es que alguna vez vivió en Massachusetts; otro es que la policía lo ha estado buscando durante muchos años; otro es que supe de su existencia por primera vez en mi juventud y otro es el hecho de que creo que todavía sigue escondido (Chisholm, 1966).

Estos son datos *acerca* de Oliver B. Garrett, y no son triviales. En general, las relaciones que existen entre las cosas del mundo en virtud de las creencias (y otros estados psicológicos) de los creyentes son relaciones acerca de las que tenemos muy buenas razones para hablar, de manera que hemos de tener *alguna* teoría o teorías capaces de aseverar que esas relaciones se conservan. Ninguna teoría metodológicamente solipsística tendrá esa capacidad, por supuesto.

La misma conclusión se hace consciente en Quine, el fundador de la literatura contemporánea acerca de la así llamada distinción *de re* y *de dicto* y le exige abandonar, de mala gana, el programa de tratar *toda* atribución de creencias (y otros estados psicológicos) como "referencialmente opaco". La no relación es la esencia del concepto de Quine acerca de la opacidad referencial; un contexto en una oración es referencialmente opaco si los símbolos que tiene lugar dentro de él no han de ser interpretados como jugando su papel normal; no son, por ejemplo, términos que están denotando lo que deno-

tan normalmente, y por tanto, no pueden estar limitados por cuantificadores. Frege sostenía un punto de vista similar, diciendo que los términos en esos contextos tenían una presencia *oblicua* y no se referían a sus denotaciones ordinarias sino a sus *significados*. Los escrúpulos ontológicos de Quine acerca de los significados fregianos y sus muchos semejantes (proposiciones, conceptos, intensiones, atributos, objetos intencionales...) lo obligan a buscar en otra parte una interpretación de la semántica de los contextos opacos (véase, por ej., Quine, 1960, pág. 151). Al final, él maneja las miriadas de los distintos predicados de creencia completa (uno para cada creencia atribuible) por analogía con la *cita directa;* tener una creencia (construida de manera no relacional) es no estar relacionado con ningún objeto u objetos en el mundo, excepto una oración cerrada. Creer es estar en un estado de otro modo no analizado captado por un predicado abultado que se distingue de los demás de su tipo por contener la inscripción de una oración que en realidad cita.

> Podríamos tratar de usar, en lugar de los objetos de la intensidad, las oraciones mismas. Aquí la condición de identidad es extrema: identidad notacional... El plan tiene sus recomendaciones. La cita no nos fallará del modo en que lo hizo la abstracción. Más aun, notoriamente opaca como es la cita es una forma vívida a la cual reducir otras construcciones opacas (1960, pág. 212; véanse también pág. 216 y Quine, 1969).

Esos predicados abultados no sirven de mucho, pero Quine ha manifestado desde hace mucho tiempo su escepticismo sobre la posibilidad de que los modismos refractarios de la intencionalidad[23] tengan algún sentido, de manera que necesita la opacidad sólo para que le proporcione una barrera de cuarentena que proteja la parte sana e intensa de una oración de la parte infectada. Aquello a lo que se renuncia mediante esta táctica. Quine cree no es nada sin lo que no se pueda vivir: "predomina una *máxima del análisis superficial: no exponga más estructura lógica de lo que parezca útil* para la deducción o cualquier otra investigación que tenga a mano. En las inmortales palabras de Adolf Meier, no se rasque donde no le pica" (1960, pág. 160). Quine evita con persistencia e ingenio la mayor parte de las exigencias aparentes para las construcciones relacionales de los modismos intencionales, pero, ante el caso que Chisholm describe, admite un picor para rascarse.

> La necesidad de una referencia recíproca desde dentro de una construcción de creencia hasta un término singular indefinido de afuera no ha de ser puesta en duda. En estos términos vea qué información urgente imparte la oración "hay alguien que creo que es un espía", en contraste con "creo que alguien es un espía" (en en sentido débil de "creo que hay espías") (1960, pág. 148).

Esto plantea entonces el problema para Quine y los autores siguientes.

[23] Se puede aceptar la tesis de Brentano ya sea como mostrando cuán indispensables son los modismos intencionales y la importancia de una ciencia de la intención autónoma, o como mostrando la falta de fundamentos de los modismos intencionales y la vacuidad de una ciencia de la intención. Mi actitud, al contrario de la de Brentano, es la segunda" (Quine, 1960, pág. 221).

"Los contextos de creencia son referencialmente opacos, por tanto carece *prima facie* de sentido cuantificar en ellos. ¿Cómo proporcionar entonces esas indispensables afirmaciones de creencias relacionales como 'hay alguien de quien Ralph cree que es un espía'?" (Quine, 1956).

A Quine se lo obliga a reconocer una distinción entre dos clases de atribución de creencia: las atribuciones *relacionales* y las *nocionales*, en sus términos, aunque otros hablan de las atribuciones *de re* y *de dicto*, y Quine reconoce que se llega a la misma cosa. (En *verdad* llega a la misma cosa en la literatura, pero si Quine hubiera querido decir lo que debería haber querido decir por "relacional" o "nocional" no habría llegado a lo mismo, como veremos.) Esto pone en movimiento la industria casera de suministrar un análisis adecuado de estas dos clases diferentes de atribución de creencia. Desgraciadamente, la forma en que Quine plantea el problema fomenta tres tipos diferentes de confusión, aunque Quine mismo no sea obviamente víctima de ninguna de ellas ni sea responsable totalmente, por supuesto, de la interpretación de sus puntos de vista que solidificó las confusiones en la literatura posterior. Primero, como Chisholm, a Quine le llama la atención una sola variedad de relación importante entre los creyentes y las cosas del mundo: casos en los que el creyente se relaciona con un individuo concreto determinado (casi exclusivamente en los ejemplos de la literatura como otra *persona*) en virtud de una creencia. Centrarse en estos casos ha llevado a una especie de ceguera institucional de la importancia de otras relaciones. Segundo, al seguir el consejo de Adolf Meier y evitar las construcciones explícitamente relacionales excepto cuando la situación lo exigía, Quine ayuda a crear la ilusión de que hay dos *tipos de creencia* diferentes, dos clases diferente de fenómenos mentales y no simplemente de dos estilos o modos diferentes de atribución de creencia. El reconocimiento de Quine de que hay veces en que uno está obligado a hacer una afirmación explícitamente relacional (y que entonces hay veces en que se puede salir del paso con una aseveración meramente nocional) se transforma en una demostración imaginaria de que hay dos clases diferentes de creencia: las relacionales y las no relacionales. En tercer lugar, si se juntan las dos primeras conclusiones, la identificación de las creencias relacionales como creencias de simples individuos determinados vuelve tentador llegar a la conclusión de que las creencias *generales* (creencias que no son intuitivamente acerca de *algo* en particular) son una variedad de creencias enteramente no relacionales. Esta conclusión subliminal ha permitido una vacilación o confusión no reconocida del status de las creencias generales para socavar proyectos por otra parte bien motivados. Trataré estas tres fuentes de confusión a su debido tiempo, demostrando cómo conspiran para creer problemas espurios y edificios de teorías que los solucionen.

Chisholm atrae nuestra atención hacia ciertos datos interesantes sobre Garrett, y Quine acusa recibo de la "información urgente" comunicada por la afirmación de que "hay alguien que Ralph cree que es un espía".[24] Pero

[24] Quine (1969) contrasta esta creencia "portentosa" con las creencias "triviales" tales como la creencia de que el espía de menor estatura es un espía; pero posteriormente en el mismo trabajo se lo lleva a un punto de vista que "anula virtualmente el contraste aparentemente vital entre [esas creencias]... al principio esto parece intolerable pero uno se acostumbra." Lo hace con una cultura adecuada que trataré de proporcionar.

considérense también otra clase de datos interesantes e importantes.

1) Mucha gente cree (erróneamente) que las víboras son viscosas.

Este es un dato acerca de la gente pero también acerca de las víboras. Es decir,

2) Las víboras son consideradas viscosas por muchos

Esta es una propiedad que tienen las víboras y es una propiedad casi tan importante como estar cubierta de escamas. Por ejemplo, es un hecho ecológico importante de las víboras que mucha gente las crea viscosas. Si no fuera así, las víboras serían ciertamente más numerosas en ciertos ambientes ecológicos de lo que son, puesto que mucha gente trata de deshacerse de las cosas que cree viscosas. La importancia ecológica de este hecho acerca de las víboras no se "reduce" a una conjunción de casos de ciertas víboras que determinada gente erróneamente cree viscosas. Muchas víboras han tenido un final prematuro (gracias a trampas o venenos para víboras, digamos) como resultado de la creencia *general* de alguien acerca de las víboras, sin siquiera haber reptado hacia una afinidad con su asesino. De manera que la relación que las víboras mantienen con cualquiera que crea en *general* que las víboras son viscosas, es una relación que tenemos motivos para querer expresar en nuestras teorías. Lo mismo sucede con la relación que cualquier víbora determinada (en virtud de su *viboridad*) mantiene con semejante creyente. He aquí otros datos interesantes elegidos para recordarnos que no toda creencia se refiere a determinadas personas.

3) Virtualmente todos creen que la nieve es fría

4) Es un hecho que muchos creen que la caridad es superior a la fe y a la esperanza

5) Muchos creen que no tener muchos amigos es peor que no tener nada de dinero

6) La democracia se valora más que la tiranía.

Lo que la cuarentena quineana de la construcción opaca tiende a ocultar de nuestra vista es que, en realidad, no se pueden hacer estas afirmaciones (en una forma que nos permita usarlas en las argumentaciones en las maneras evidentes) a menos que se las pueda hacer de un modo que permita que se expresen las relaciones explícitas. Nos ayudará considerar con más detalle un caso singular.

Sam es un iraní que vive en California y Herb cree que todos los iraníes de California deberían ser deportados inmediatamente, pero no tiene idea de quién es Sam y en realidad ni siquiera tiene indicios de su existencia, aunque sabe, por supuesto, que hay iraníes en California. Suponiendo que Herb sea una autoridad o simplemente un ciudadano influyente, es una creencia que Sam, que disfruta de California, lamentaría. Para Sam un mundo en el que la gente tiene esta creencia es peor que un mundo en el que nadie la tiene. Digamos que Sam *está en peligro* por esta creencia de Herb. ¿Quién más corre peligro? Todos los iraníes que viven en California. Supongamos que Herb cree también que todos los fumadores de marihuana deberían ser azotados en público. ¿Está creencia también pone en peligro a Sam? Por supuesto, depende de si Sam es un fumador de marihuana. Ahora bien, algo se deduce de

7) Sam es un iraní que vive en California y

8) Herb cree que todos los iraníes que viven en California deberían ser deportados inmediatamente

no es una consecuencia de 7) y

9) cree que todos los fumadores de marihuana deberían ser azotados en público.

El resultado es algo que permite la conclusión de que Sam está en peligro debido a la creencia de Herb citada en 8), pero podemos coincidir en que lo que sigue *no* lo es

10) Herb cree *acerca de Sam* que debería ser deportado de inmediato.

Es decir que ninguna creencia del tipo que impresiona a Quine y Chisholm se deduce de 7) y 8). Más bien, estamos buscando algo más parecido a

11) Sam es un miembro del grupo de iraníes de California, y es *sobre ese grupo* que Herb cree que todos sus miembros deberían ser deportados de inmediato.

Algunos podrían estremecerse ante la idea de tener bastante *relación* con un grupo como para tener creencias *de re* respecto a él, pero en ningún caso los grupos harán el trabajo. Sam, sabiendo el riesgo que corre puede desear librarse de él, por ejemplo, yéndose de California o cambiando la creencia de Herb. Si se fuera de California cambiaría el número de miembros del grupo pertinente —modificaría el grupo importante— pero por cierto que no alteraría la creencia de Herb. En pocas palabras, no es ser un miembro del grupo lo que pone en peligro a Sam sino tener un atributo.

12) Hay un atributo (la calidad de ser iraní californiano) tal que Sam lo tiene y Herb cree *acerca de él* que cualquiera que lo tenga debería ser deportado de inmediato ¡No se trata sólo de cuantificar en los atributos sino también por encima de ellos! Se podría intentar suavizar el golpe ontológico jugueteando con circunloquios según los lineamientos de

13) $(\exists\alpha)$ $(\exists c)$ [α denota c & Sam es miembro de c y Herb cree cierto (x) (x es miembro de $\alpha \supset x$ debería ser deportado inmediatamente)]

pero no funcionarán sin condiciones *ad hoc* de distintos tipos. Si hemos de tomar en serio lo que se dice acerca de las creencias, podríamos del mismo modo permitir que todo sucediera como lo queremos y pemitir la cuantificación por encima de atributos y relaciones, así como los individuos y las clases en todas las posiciones dentro de los contextos de creencia.[25] Véase Wallace (1972) para una propuesta de carácter similar.

La defensa de este curso ha estado delante de nosotros desde el principio, puesto que es una implicación de las argumentaciones putnamianas de que hasta las actitudes proposicionales generales, si son auténticamente *proposicionales* en cuanto a cumplir con las dos primeras condiciones de Frege deben implicar una relación entre el creyente y las cosas del mundo. *Mi* creencia de que todas las ballenas son mamíferos, *se refiere a las ballenas*; la

[25] Quine abandona explícitamente este rumbo mientras todavía espera captar cualquier inferencia útil e importante que hubiera (1960 pág. 221; 1969) para mayores argumentos a favor de considerar algunos casos de estados psicológicos como actitudes *de re* hacia las propiedades y relaciones, véase Aquila 1977, especialmente págs. 84-92.

creencia contrafigura de mi *Doppelgänger* podría no serlo (si su Planeta Tierra Gemelo tuviera peces grandes llamados "ballenas", por ejemplo). Son las ballenas las que creo que son mamíferos y no se podría realmente decir que yo creo *que* las ballenas son mamíferos, a menos que se pudiera decir también *acerca* de las ballenas que las creo mamíferos.

¿Deberíamos decir que las creencias enteramente generales son acerca de algo? Se podría leer nuestra condición fregiana (*b*) como exigiéndolo, pero hay modos de negarlo (véase, por ejemplo, "What do General Propositions Refer to?" y "Oratio Obliqua" en Prior, 1976). Por ejemplo, se puede notar que la forma lógica de las creencias generales, tales como la creencia de que todas las ballenas son mamíferos es (x) $(Fx \supset Gx)$, que dice, en efecto, que cada cosa es de tal manera que si es una ballena es un mamífero. Semejante afirmación se refiere tanto a las coles y los reyes como a las ballenas. Al referirse a casi todo, no se refiere a nada (véase Goodman, 1961; Ullian y Goodman, 1977; Donnellan, 1974). Esto no ayuda mucho a acallar la intuición de que cuando alguien cree que las ballenas son peces está *equivocado acerca de las ballenas*, no equivocado acerca de todo. He aquí otro desafío molesto: si las creencias generales se refieren siempre a las cosas mencionadas en su expresión, ¿cuál es la creencia de que no hay unicornios? ¿Unicornios? No hay ninguno. Si estamos preparados, como he exhortado a que lo estemos, para cuantificar por encima de los atributos, podemos decir que esta creencia se refiere a la *unicorniedad*, en el sentido de que no está ejemplificada completamente en ninguna parte.[26] Si hubiera unicornios, la creencia sería una creencia falsa acerca de los unicornios, exactamente del mismo modo en que la creencia de que no existen las ballenas azules es (en la actualidad) una creencia falsa acerca de las ballenas azules. Mientras señalamos un gran grupo de ballenas podríamos decir: Tom no cree en la existencia de estas criaturas.

En todo caso este tira y afloja acerca del "acerca" no me parece una estrategia provechosa a seguir (véase Donnellan, 1974) y sin embargo es necesario decir algo más de este término conflictivo pero prácticamente indispensable, mencionado una vez y utilizado dos veces en esta misma oración. "El término 'acerca' es notoriamente vago que notoriamente no debe ser confundido con 'denota'" (Burge, 1978, pág. 128). Tal vez haya cierto sentido de "acerca" que debe distinguirse cuidadosamente de "denota" —gran parte de la literatura acerca de la creencia *de re* es después de todo una indagación para explicar un sentido tan *fuerte* de "acerca— pero hay innegablemente otro sentido más débil (y de hecho mucho más claro) de "acerca" cuya esencia es la denotación. Supongamos que creo que el espía de menor estatura es mujer. (No piensa en nadie en particular, como se dice, pero parece una buena apuesta.) Ahora bien, puesto que no pienso en nadie en especial, como se dice, puede haber, por cierto, un sentido en el que mi creencia no es *acerca de* nadie. Sin embargo mi creencia es verdadera o falsa. Lo que sea depende del género de cierta persona real, el espía de menor estatura quienquiera que sea. Esa persona es la verificadora o falsificadora de mi creencia, la que

[26] Los puntos de vista de Kripke acerca de los unicornios exigen que no sólo no haya unicornios sino que no podría haber ninguno. ¿Podría existir entonces el atributo de unicornidad? Véase el Apéndice de Kripke, 1972, págs. 763-69.

satisface la descripción definida que usé para expresar la creencia. Existe *esa* relación entre mí y el espía de menor estatura en virtud de mi creencia, y por más despreciable y carente de interés que alguien pudiera encontrar esa relación, es una relación que tengo precisamente con una persona en virtud de mi creencia y prácticamente no podríamos darle un nombre mejor que acerquidad, acerquidad *débil* podríamos llamarla permitiendo el posible descubrimiento posterior de tipos más fuertes y más interesantes de acerquidad. Supongamos que Rosa Klebb es la espía de menor estatura. Sería entonces enganoso decir que yo pensaba en ella cuando se me ocurrió que la espía de menor estatura era probablemente una mujer, pero en este sentido débil de "acerca" sería sin embargo cierto.[27]

La segunda confusión de la cual Quine mismo es, aparentemente, inocente está fomentada por ciertas instrucciones erróneas en sus "Quantifiers and Propositional Attitudes" (1956), el trabajo más importante en plantearle problemas a la literatura posteror. Quine ilustra la diferencia entre dos de los sentidos de "cree" con un par de ejemplos.

> Considere los sentidos relacionales y nocionales de creer en espías:
> 14) ($\exists x$) (Ralph cree que x es un espía)
> 15) Ralph cree que ($\exists x$) (x es un espía)
> ...La diferencia es amplia; por cierto, si Ralph es como la mayoría de nosotros 15) es verdad y 14) es falso (pág. 184).(He cambiado la numeración de los ejemplos de Quine para adaptarlos a mi secuencia.)

Los ejemplos por cierto ilustran dos estilos diferentes de atribución: la 14) de Quine está ontológicamente comprometida mientras que 15) es reticente; pero también ilustran tipos muy diferentes de estados psicológicos. 14) es acerca de lo que podemos llamar una creencia *específica*, mientras que 15) es acerca de una creencia *general* (véase Searle, 1979). Ha resultado irresistible para muchos extraer la conclusión injustificada de que 14) y 15) ilustran respectivamente una clase relacional y una clase nocional de creencia. Lo que la gente desconoce es la posibilidad de que *ambas* creencias se pueden atribuir en los dos estilos *relacionales* y *nocionales* de la atribución. Por una parte, alguna otra afirmación relacional [distinta a que 14) puede ser verdad en virtud de 15)], por ejemplo,

16) ($\exists x$) (x es espionaje y Ralph cree que x es una abstracción ejemplificada concretamente),

[27] De este modo también sería verdad que *usted* está hablando *acerca de ella* (en este sentido débil) cuando asevera que yo pensaba acerca del espía de menor estatura. Por lo tanto debo disentir con Kripke: "Si una descripción está inserta en un contexto (de intensión) ***de dicto*** no se puede decir que estamos hablando *acerca de la cosa descrita, ya que qua* su respuesta a la descripción o *qua* cualquier otra cosa. Tomado ***de dicto***, "Jones cree que la debutante más rica de Dubuque se casará con él', puede ser afirmado por alguien que piensa (supongamos erróneamente) que no hay debutantes en Dubuque. Entonces, por cierto, él *de ninguna manera* [el subrayado es mío] está hablando acerca de la debutante más rica ni siquiera 'en forma atributiva'" (1977). Imagínese a Debby, que sabe muy bien que es la debutante más rica de Dubuque alcanzando a oír este comentario. Podría decir: "Podrás no saberlo, hombre, pero estás hablando acerca de mí y me has hecho reír mucho. Puesto que aunque Jones no me conozca, yo sé quién es él: un pobrecito trepador social, y no tiene ninguna chance".

y por otra parte, tal vez haya otras lecturas meramente nocionales distintas de 15) que surgen de 14). Muchos lo han creído. Todos los que han tratado de aislar "un componente *de dicto*" en la creencia *de re* lo han creído, en efecto, aunque sus esfuerzos han sido típicamente confundidos por los malentendidos acerca de qué podrían ser las creencias *de dicto*. Al contrastar 14) con 15) Quine compara manzanas y naranjas, un desajuste oculto por la posibilidad espuria de la idea de que el cambio lexicalmente simple efectuado al cambiar de lugar el cuantificador nos lleva de un lado a otro entre una afirmación relacional y su contrafigura nocional más próxima. La verdadera contrafigura nocional de la relacional

17) ($\exists x$) (Es creído por Tommy que x llenó su calcetín con juguetes)

no es

18) Tommy cree que ($\exists x$) (x llenó su calcetín con juguetes),

sino algo para lo cual no tenemos una expresión formal, aunque su fuerza intencional se puede expresar mediante una cita sensacionalista:

19) Tommy cree "acerca de Papá Noel" que le ha llenado el calcetín con juguetes.[28]

El mundo nocional de Tommy está habitado por Papá Noel, un hecho que queremos expresar al describir el estado de creencia presente de Tommy sin comprometernos acerca de la existencia de Papá Noel, que es lo que haría la afirmación relacional (17). Este es el trabajo para el que incorporamos la discusión acerca del mundo nocional. De manera semejante, pero moviéndonos en la dirección opuesta, cuando queremos distinguir mi creencia general de que toda agua es H_2O de su gemelo nocional en mi *Doppelgänger*, tendremos que decir, relacionalmente, que *agua* es lo que yo creo que es H_2O.

Esto no intenta demostrar por sí mismo que hay dos tipos diferentes importantes de creencia, o más que dos, sino sólo que cierto alineamiento conocido de los temas es un desalineamiento. Un síntoma ulterior especialmente insidioso de este desalineamiento, es el mito de la no relacionalidad de las creencias *"de dicto* puras". (¡Qué podría ser más evidente se podría decir: si las creencias *de re* con creencias relacionales, entonces las creencias *de dicto* deben ser creencias no relacionales!) La posición ortodoxa es sucintamente resumida —no defendida— por Sosa (1970):

20) "La creencia *de dicto* es la creencia de que cierto *dictum* (o proposición) es cierta, mientras que la creencia *de re* es la creencia acerca de una *res* determinada (o cosa) que tiene cierta propiedad."

Podríamos detenernos a preguntarnos si la siguiente definición podría ser un sustituto aceptable:

21) La creencia *de dicto* es la creencia *acerca de* cierto *dictum* (o proposición) es verdad, mientras que la creencia *de re* es la creencia *de que* determinada *res* (o cosa) tiene cierta propiedad,

y si no por qué no. Pero en lugar de proseguir ese interrogante quiero centrarme en cambio en la unión de las palabras "dictum (o proposición)", y lo que se puede conjurar a partir de ellas. El latín es refinadamente ambiguo: significa *lo que se dice*, ¿pero significa eso *lo que se pronuncia* (las pa-

[28] Sellars (1974) discute la fórmula: "Jones cree acerca de alguien (que puede o no ser real) que es sabio".

labras mismas) o *lo que se expresa* (lo que las palabras se utilizan para afirmar)? La *OED* nos dice que un dictum es "un dicho", que brinda una versión más de la equivocación. Vimos al comienzo que estaban aquellos que consideraban las proposiciones como cosas tipo oración y aquellos que las veían más bien como los significados (abstractos) de *(inter alia)* de cosas parecidas a oraciones. Estas son concepciones fundamentalmente diferentes como hemos visto, pero con frecuencia se las mantiene convenientemente no distinguibles en la literatura acerca de *de re* y *de dicto*. Esto permite la influencia secreta de una figura incoherente: las proposiciones como *dichos mentales*, sucesos enteramente internos que deben sus identidades a nada más que sus propiedades intrínsecas y, por tanto, son enteramente no relacionales.[29] Al mismo tiempo, estos dichos mentales no son meras *oraciones*, meros objetos sintácticos sino *proposiciones*. Se alienta la idea de que una atribución de actitud proposicional *de dicto* es una determinación de contenido completamente interna o metodológicamente solipsística que es independiente de cómo el creyente está situado en el mundo. La metafísica subconsciente de esta seguridad acerca de las formulaciones *"de dicto"* como especificadoras del contenido, es la idea de que de algún modo la aceptación mental de una *oración* causa la *proposición* que la oración *expresa* como *creída*.[30] (Los detalles del mecanismo de adopción se dejan para el psicólogo. Lo que me resulta levemente sorprendente es la aparente voluntad de muchos psicólogos y teóricos de *AI* para aceptar este planteamiento de su tarea sin un recelo aparente.)

Nadie es más claro que Quine acerca de la diferencia entre una oración y una proposición, y sin embargo, el truco con el que intenta hacer desaparecer las proposiciones de la escena y salir del paso solamente con oraciones puede haber contribuido a la confusión. En *Word and Object* (1960), después de presentar los distintos problemas de los contextos de creencia y juguetear con un sistema desarrollado de proposiciones, atributos y relaciones en intensión para manejarlas, Quine demuestra cómo renunciar a estas "criaturas de la oscuridad" en favor de los predicados abultados de la cita directa de las oraciones. Con anterioridad (1956) había hecho una curiosa defensa de esta jugada:

> Cómo, dónde y sobre qué fundamento trazar un límite entre aquellos que creen o desean o luchan que *p*, y aquellos que no creen o desean o luchan completamente que *p* es un asunto innegablemente vago y oscuro. Sin embargo, si alguien en verdad aprueba que se hable de la creencia de una proposición y que se

[29] En este sentido, los dichos mentales se ven semejantes a Qualia: características *aparentemente* intrínsecas de las mentes a ser contrastadas con características relacionales funcionalmente caracterizables. Yo discuto la incoherencia de esta imagen de Qualia en "Quining Qualia" (de próxima aparición d). Para una expresión clara del punto de vista sospechoso, considere a Burge (1977 pág. 345: "Desde un punto de vista semántico una creencia *de dicto* es una creencia en la que el creyente está relacionado sólo con una proposición completamente expresada *(dictum)*"].

[30] Kaplan (1980) hace un uso cautivadoramente explícito de esta imagen al proponer "usar la diferencia entre el discurso directo e indirecto para conjugar la diferencia entre carácter y contenido" (de pensamientos o creencias ahora, no de oraciones); esto a pesar de su concesión igualmente cautivadora de que "no existe una verdadera sintaxis del lenguaje del pensamiento".

hable a su vez de una proposición como expresada por una oración, ciertamente *entonces no puede objetar nuestra reformulación semántica* "*w* cree que S es verdad" en cualquier terreno especial de oscuridad; puesto que "*w* cree que S es verdad" es definible explícitamente en *sus* términos como "*w* cree en la proposición expresada por S" (pág. 192-193).

Esto nos explica cómo el creyente en proposiciones, atributos y cosas por el estilo ha de encontrarle sentido a los nuevos predicados de Quine, pero no nos explica cómo Quine les encuentra sentido. Nunca nos lo dice, pero dada la *"huida de la intensión"* por la cual aboga para todos nosotros, se puede suponer que no cree que realmente tenga mucho sentido. No son más que un recurso para pasar por alto un aspecto malo del material que otra gente insiste en insertar en el cuerpo principal de la elocución seria. La función de las paráfrasis de Quine es permitirle al funcionario público que traduce el conjunto a la "notación canónica" que llegue "a salvo" al final de la oración y pase a la oración siguiente confiado en que ha hecho *por lo menos* los suficientes predicados nuevos como para asegurar que nunca use el mismo predicado para traducir dos afirmaciones de la actitud intencional de intención diferente. Si alguna vez se llega a encontrar el sentido de esas afirmaciones se las puede recuperar sin pérdida de claridad del congelamiento canónico. Entretanto, los predicados de Quine son excesivamente nocionales, pero también inertes. El problema de esto surge en tres citas:

A) En el... sentido opaco "quiere" no es un término relativo que relacione a la gente con algo, concreto o abstracto, real o ideal (1960, págs. 155-56).

B) Si la creencia se la toma opacamente entonces (*Tom cree que Cicerón denunció a Catilina*) no relaciona expresamente a Tom con ningún hombre (página 145).

C) 1) "Tulio era romano" es trocaico.

2) El comisionado está buscando al presidente de la junta del hospital.

El ejemplo 2), aun cuando no está tomado en el sentido puramente referencial, difiere de 1) en que todavía parece tener mucho más predicamento sobre el presidente de la junta del hospital, por más decano que él sea, que el que tiene 1) sobre Tulio. De ahí mi frase cautelosa "no puramente referencial" destinada a ser aplicada a todos los casos semejantes y a afirmar que no hay ninguna diferencia entre ellas. Si omito el adverbio, la razón será la brevedad (pág. 142).

A) es inflexible: no hay *nada* relacional en el sentido opaco de "querer"; en B) la afirmación opaca no relaciona *expresamente* a Tom con nadie; en C) se reconoce que el ejemplo 2) "tiene una relación" como una persona determinada, y por tanto debe ser tomado como algo referencial, pero no puramente referencial. En realidad, en todos estos casos hay una evidencia clara de lo que podríamos llamar *referencialidad impura*. Desear un velero "en el sentido nocional" no relaciona a alguien preeminentemente con ningún velero en especial, pero el alivio del deseo del "velerismo", como lo llama Quine, tiene, no obstante, relación con los veleros; si nadie deseara aliviarse del velerismo, los veleros no serían ni tan numerosos ni tan caros como son. No se puede creer *que* Cicerón denunció a Catilina sin creer *acerca* de Cicerón que denunció a Catilina. Y es un hecho acerca del presidente de la junta del hospi-

tal que está siendo buscado por el comisionado, en virtud de la verdad de 2).

Esta referencialidad impura fue observada por varios autores y Quine mismo proporciona el paradigma para el análisis posterior con su ejemplo.

22) A Giorgione se lo llamó así a causa de su talla (1960, pág. 153), en el que, como lo nota Quine, "Giorgione" cumple una función doble; la expansión clara de 22) es

23) A Giorgione se lo llamó "Giorgione" debido a su talla.

Castañeda (1967), Kiteley (1968), Loar, (1972) y Hornsby (1977) desarrollan el tema del papel *normalmente* dual de los términos singulares dentro de las cláusulas de oraciones de la actitud proposicional, para explicar, por ejemplo, el papel aparente del pronombre en oraciones tales como

24) Michael cree que ese hombre enmascarado es un diplomático, pero él obviamente no lo es (Loar, 1972, pág. 49).

Esto corrige parte de la historia, pero no percibe la extensión de la doctrina de los términos singulares a los generales. Cuando cuantificamos por encima de los atributos, como en 12) y en 16), completamos el cuadro. Un término general dentro de una cláusula proposicional en una oración de la actitud proposicional juega normalmente un papel dual, tal como lo sugieren oraciones tales como

25) Herb piensa que todos los iraníes de California tendría que ser deportados, pero ninguno de ellos tiene nada que temer de parte de él. [31]

Una vez que distinguimos las atribuciones nocionales de las relacionales, y distinguimos *esa* distinción de la distinción entre creencias específicas y generales, podemos ver que muchas de las afirmaciones que se han hecho en el esfuerzo por caracterizar la "creencia *de re*" se aplican a todas las categorías y no distinguen una clase especial de creencia. Consideremos primero la distinción entre las atribuciones nocionales y relacionales que surge cuando abordamos el problema de relacionar los estados psicológicos limitados de un individuo con sus estados proposicionales amplios. Comenzamos con el mito de que hemos determinado el mundo nocional del sujeto, y que ahora tenemos que alinear sus actitudes nocionales con un conjunto de actitudes proposicionales. Aquí se trata de encontrar al más apto. Como hemos visto, en los casos de sujetos gravemente confundidos, puede no haber ninguno apto. En el caso de un sujeto irracional, no habrá ningún apto perfecto porque no hay *ningún mundo posible* en el que se pueda ubicar al sujeto de manera feliz. En el caso de un sujeto muy mal informado, habrá simplemente un área de desajuste con el mundo *real*. Por ejemplo, el niño que "cree en Papá Noel" tiene a Papá Noel en su mundo nocional, y nadie en el mundo real es una contrafigura adecuada, de manera que algunas de las actitudes nocionales del niño acerca de Papá Noel no se pueden cambiar por ***ninguna*** actitud proposicional en absoluto. Son como símbolos de oraciones cuyo carácter es tal que en el contexto en el que ocurren no determinan *ningún* contenido. Como alega Donnellan cuando un niño así *dice*: "Esta noche viene Papá Noel", lo

[31] Cuando Burge (1977) duda de la existencia de las creencias *de dicto* "puras" creo que se lo puede leer mejor como expresando un reconocimiento de este punto. Véase también Field (1978, pág. 21-23).

que dice no expresa ninguna proposición (1974, pág. 234). Podemos agregar: el estado psicológico del niño en ese momento es una actitud nocional que no determina ninguna actitud proposicional. McDowell (1977) y Field (1978) también endosan diferentes versiones de este alegato, al que se lo puede hacer aparecer como extravagante para cualquiera que insista en considerar a las creencias como (nada más que) actitudes proposicionales. Puesto que en ese caso hay que decir que nadie tiene creencias acerca de Papá Noel, ni podría tenerlas; ¡algunas personas simplemente creen que las tienen! Quienes creen en Papá Noel tienen muchas actitudes proposicionales, por supuesto, pero también tienen algunos estados psicológicos que no son en absoluto ninguna actitud proposicional. ¿Qué tienen en su lugar? (Véase Blackburn, 1979). Actitudes nocionales. La atribución de actitudes nocionales al niño que cree en Papá Noel nos proporcionará toda la comprensión y la eficacia teórica que necesitamos para explicar la conducta del niño. Por ejemplo, podremos obtener actitudes proposicionales auténticas de las actitudes meramente nocionales del niño cuando el mundo ocasionalmente lo obligue. (House, inédito.) Tommy cree que el hombre de la falsa barba que está en la gran tienda le traerá regalos porque cree que el hombre es Papá Noel. Esa es la razón por la cual Tommy le *dice* lo que quiere.

En otras ocasiones, por ej., en la Pizzería de Shakey, hay una confusión de opulencia: demasiados objetos del mundo real como *relata* candidatos cuando nos volvemos relacionales y permutamos afirmaciones de la actitud nocional por afirmaciones de la actitud proposicional. Esta posibilidad no está restringida a creencias específicas acerca de individuos. Las discusiones de Putnam (1975a) de los términos bondadosos naturales revelan que el mismo problema surge cuando debemos decidir cuál proposición general, en la que alguien cree cuando cree en algo la expresaría con las palabras "Todos los olmos son caducos", cuando todos sabemos que él no distingue un olmo de un haya. En un caso así, las diferentes propiedades o atributos son las *relata* candidatas para las creencias.

La diferencia entre creencias generales y específicas todavía no se ha aclarado, y es sólo aparente, pero entretanto podemos observar que cuando nos hacemos relacionales, permutaremos afirmaciones específicas de la actitud nocional por afirmaciones específicas relacionales, es decir, de la actitud proposicional, y afirmaciones generales de la actitud nocional por afirmaciones generales relacionales, es decir, de la actitud proposicional. La creencia trivial de Ralph citada nocionalmente en 15) también se puede citar relacionalmente como 16), mientras que la creencia portentosa de Ralph, citada relacionalmente en 14), también podría ser citada nocionalmente si hubiera un motivo: por ejemplo, si Ralph hubiera sido llevado con engaños a su estado de urgencia por bromistas que lo hubieran convencido de la existencia de un hombre que nunca existió. Las condiciones necesarias para mantener una atribución relacional son completamente independientes de la distinción entre las actitudes específicas y generales. Esto se puede comprobar si incorporamos ejemplos de creencias generales en las conocidas arenas de la argumentación.

26) Tom cree que las ballenas son mamíferos.

Esta es una creencia general de Tom, si alguna lo es, pero con anterioridad sugerimos que no se puede decir en verdad que Tom cree *que* las ballenas son mamíferos a menos que estemos preparados para decir también *acerca de las ballenas* que Tom las cree mamíferos. Formalmente, tal afirmación de creencia permitiría el cuantificar de la manera siguiente:

27) $(\exists x)$ (x = ballenidad & Tom cree acerca de x que cualquier cosa que explique esa abstracción concretamente es un mamífero)

Entonces, ¿no debe tener Tom una noción *vívida* de la ballenidad? (Kaplan, 1968).

Contrastemos 26) con

28) Bill cree que los mamíferos más grandes son mamíferos.

Si todo lo que Bill sabe acerca de las ballenas es que son los mamíferos más grandes, tiene una noción muy poco vívida de la ballenidad. La mayor parte de nosotros tenemos una afinidad mayor con las ballenas, gracias a los medios, pero consideremos 29) que Tom cree que las ballenas son mamíferos.

Hay por cierto una influencia intuitiva en contra de decir que Tom cree mamíferos a las ballenas, puesto que Tom, como la mayor parte de nosotros está cognitivamente muy lejos de las ballenas. Se podía acercar aprendiendo (por medio del diccionario) que las ballenas son de la orden Sirenia, de grandes hervíboros acuáticos que se pueden distinguir de los manatíes por la cola bilobulada. Ahora bien, ¿está Tom lo suficientemente informado como para creer *de* las ballenas que son mamíferos ¿Lo está usted? Una cosa es creer, como podría hacerlo un alemán monolingüe, que la oración "Las ballenas son mamíferos" es cierta; eso no cuenta para nada para creer que las ballenas son mamíferos. Otra cosa es saber inglés y creer que la oración "Las ballenas son mamíferos" es cierta y por lo tanto creer (en un sentido muy mínimo) que las ballenas son mamíferos. Ese es aproximadamente el estado en que la mayor parte de nosotros estamos.[32] Si se ha leído un libro acerca de las ballenas o se ha visto una película o —mejor aun— se ha visto una ballena en un zoológico o se ha tenido una ballena como mascota, se está mucho más seguro en el status como creyente de que las ballenas son mamíferos. No se trata de las condiciones para creer que alguna ballena en especial sea un mamífero, sino en creer en general que (todos) las ballenas son mamíferos.[33]

Inicialmente Kaplan propuso las nociones vívidas como condición de la

[32] Hornsby (1977) discute el caso de "Jones, un individuo sin educación que ha... encontrado 'Quine' en una lista de nombres de filósofos. No sabe absolutamente nada de ese hombre pero llega a creer simplemente que no podría afirmar alguna verdad con las palabras 'Quine es un filósofo'. En circunstancias como éstas, una lectura relacional... no puede ser correcta. Pero en estas circunstancias Jones en realidad no cree más que en las propiedades de llamarse 'Quine' y ser filósofo coinciden en alguna parte" (pág. 47). Sin embargo la última oración es *una* lectura relacional y según las circunstancias otras lecturas relacionales son posibles: por ej., la propiedad de ser la persona llamada "Quine" a quien los autores de esta lista tuvieron intención de incluir en la lista se explica en forma concreta por medio de alguien que también explica en forma concreta la propiedad de ser filósofo.

[33] Richmond Thomson, durante una conversación, sugiere que los problemas planteados por estos casos [por 1), 2), 25), 29)] son en realidad problemas acerca de la lógica de los plurales. El distinguiría creer que las ballenas son mamíferos de creer que todas las ballenas son mamíferos. Tal vez debamos hacer esta diferencia —y de ser así, es una lástima, puesto que Thomason

auténtica creencia *de re*, pero sostuvo que no había ningún umbral de vivacidad por encima del cual podría elevarse una noción para que uno creyera *de re* con ella. Desde entonces, ha abandonado la vivacidad como una condición de la creencia *de re*, ¿pero por qué fue la idea tan atractiva al principio? Porque la vivacidad, o lo que es casi igual a ella tal como Kaplan la caracterizó (1968) es una condición acerca de *toda* creencia; uno debe estar sumamente informado acerca del mundo en general, sus ocupantes y propiedades e íntimamente conectado con ellos para que se le puedan adjudicar creencias con alguna certeza. Todas las "nociones" deben ser algo vívidas. La idea de que algunas creencias, las creencias *de dicto* no tienen ninguna exigencia de vivacidad, es un síntoma del punto de vista subliminal de las creencias *de dicto* como meros dichos mentales, no tan diferentes del estado del alemán monolingüe cuando de algún modo ha adquirido "Las ballenas son mamíferos" como una oración cierta.

No obstante, otro tema de la literatura se ocupa de la admisibilidad o inadmisibilidad de la sustitución de expresiones co-indicativas dentro de las cláusulas proposicionales de las atribuciones de creencia. Se dice que cuando uno está atribuyendo una creencia *de re*, la sustitución es permisible; de otro modo, no. Pero si todas las creencias generales tienen rendimientos relacionales (en los que se las ve como creencias *de* ciertos atributos, a la manera de 12) y 27) ¿admiten éstas la sustitución? Si, con las mismas clases de restricciones pragmáticas que aseguren la referencia sin apariencias falsas que se han observado en la literatura de las creencias acerca de los individuos (Sosa 1970; Boer y Lycan, 1975; Hornsby, 1977; Searle, 1979; por nombrar sólo algunos).

30) Mi esposa quiere que le compre un suéter del color de su camisa.

31) Su abuelo pensaba que los niños que se portaban como usted se está portando, deberían ser mandados a la cama sin cenar

y por supuesto el famoso

32) Yo creía que su yate era más largo de lo que es

de Russell.

En 30) y 31) nos hemos referido a los atributos por medio de la descripción (véase Aquila, 1977, pág. 91). Algunas veces la referencia a los *objetos* está asegurada por algo parecido a la ostentación directa, a menudo con la ayuda de demostrativos, y Donnellan (1966, 1968, 1970, 1974), Kaplan (1978, 1980) y otros han afirmado que hay una diferencia fundamental entre la referencia *directa* y la referencia de algún modo indirecta que se obtiene por la vía de la descripción definida. Posponiendo una vez más la consideración de los méritos de estas afirmaciones, observemos que la supuesta distinción no marca ninguna característica especial de la creencia específica acerca de determinados objetos. Podemos decir, en un acto de referencia directa

33) *Esta* es la torre Eiffel y *ése* es el Sena.

También podemos decir con el mismo estado de ánimo:

---

afirma que todavía nadie ha ideado una explicación de los plurales libre de problemas— pero aun si hacemos esa distinción creo que nada excepto la naturalidad se pierde si damos una nueva forma a mis ejemplos explícitamente como los ejemplos de creer que todas las víboras son viscosas, todas las ballenas son mamíferos y así sucesivamente.

34)Un cuerno inglés suena *así* mientras que un oboe suena *así*.

35) Para lograr un buen vibrato, haga *esto* (véase Jackendoff, 1985).

Un francés dice que algo tiene un cierto *je ne sais quoi*; lo que quiere decir es que *il sait* perfectamente bien *quoi*, el no puede *say quoi*, porque la propiedad en cuestión es un *quale* como el gusto del ananá o el modo como luces esta noche (véase Dennett, de próxima aparición d). El puede predicar la propiedad de algo sólo mediante una referencia identificatoria a la propiedad por medio de una descripción tan definida, y si esto ocurre dentro de un contexto de intensión a menudo debe leérselo en forma trasparente, como en 30) y 31) o

36) Rubens creía que las mujeres que tienen el aspecto de usted son muy hermosas.

Con la relacionalidad, la vivacidad, la sustitutabilidad y la referencia demostrativa directa colocadas detrás de nosotros podemos volver a la pregunta postergada:

¿Cuál es la diferencia entre la creencia general y la específica? Deberíamos poder deducir la diferencia independientemente de las atribuciones auténticamente relacionales, por tanto al nivel de las atribuciones de la actitud nocional, puesto que queremos distinguir los estados mentales específicos acerca de Papá Noel, de otros estados mentales generales, por ejemplo.

Hay dos puntos de vista opuestos en la literatura filósofica que deben ser desenredados de su compromiso habitual con los usos referenciales y atributivos de descripciones definidas en los actos de discursos públicos y con los problemas que surgen de las tentativas por captar una teoría de referencia específicamente *causal*. A una se la puede llamar el punto de vista de la Descripción Definida, y a la otra, su negación, que toma distintas formas, se la puede llamar evasivamente, el punto de vista de la Referencia Especial. En cuanto al primer punto de vista, la única distinción a observar es que entre creer que *todas* las *Fs* son *Gs* por un lado (las creencias generales), y creer que *sólo una F* es *G*, por otro (creencia específica), donde esas creencias específicas son consideradas como habiendo sido captadas adecuadamente por la teoría de Russell de las descripciones definidas. El segundo punto de vista, aunque no niega la existencia de esa direrencia en forma lógica, insiste en que aun las creencias a las que el primer punto de vista llama específicas, las creencias expresadas por las descripciones russellianas definidas son apropiadamente generales, aunque hay una categoría ulterior de creencias específicas de verdad, que son más intensamente *acerca* de sus objetos porque los eligen por alguna clase de referencia *directa*, no dirimida por (el sentido de) descripciones de ninguna clase.

Para la gran frustración de cualquiera que intente dirigir el punto como lo haría digamos, un psicólogo o un teórico AT, ninguno de los dos puntos de vista tal como están típicamente expresados en la literatura filosófica establece mucho contacto plausible con lo que debe ser, al fin, el punto empírico: qué es, *literalmente* en las cabezas de los creyentes, lo que hace de un estado psicológico una creencia acerca de un objeto determinado. El punto de vista de la Descripción Definida, si se toma literalmente su valor nominal, sería un oracionalismo ridículo, reconociendo que, sus adherentes gesticulan en la dirección de los mecanismos de reconocimiento, los procedimientos y las pruebas

juiciosas como medios de captar el efecto de las descripciones definidas en términos más psicológicamente realistas (si sólo fuera porque es más vaga y por lo tanto menos irreal). Por otra parte, los críticos de todas esas propuestas tienen sus propios gestos que hacer en dirección de intimidades causales no especificadas y rutas genealógicas entre los estados psicológicos y sus objetos verdaderos. Puesto que no se puede lograr ningún progreso a esos niveles metafóricos, tal vez sea mejor retroceder hacia los términos filosóficos tradicionales (pasando por alto la psicología por un rato) para ver cuál podría ser el punto.

Se puede observar en favor del punto de vista de la Descripción Definida que una descripción definida russelliana logra, por cierto, escoger un objeto específico del campo de acción del discurso cuando todo va bien. El punto de vista de la Referencia Especial puede sostener, en contra de esto, que la expansión russelliana de una descripción definida revela su generalidad disfrazada: según el análisis russelliano, afirmar que *el hombre que mató a Smith es un insano* es afirmar que cierta persona *x* tiene cierta propiedad (alguna *F* es *G*): alguna *x* tiene la propiedad de ver idéntica a cualquier prsona que haya matado a Smith y sea insana. Esta generalidad se revela también ante la intuición (se afirma) por el test "quienquiera": "el hombre que mató a Smith", si significa "*quienquiera* que haya matado a Smith" (como debe ser, según el análisis russelliano) no hace ninguna referencia específica. Intuitivamente, "quienquiera que haya matado a Smith" no *señala* a nadie en especial, sino más bien *arroja una red* en el campo de acción. Una creencia específica auténtica señala directamente a su objeto, que se identifica no por ser el único portador de alguna propiedad, sino por ser... el objeto de la creencia. ¿Por qué tendríamos que suponer que existen creencias así? Aquí hay enredadas dos motivaciones diferentes, una metafísica, que tiene que ver con el esencialismo, y la otra psicológica, que tiene que ver con las diferencias en el estado psicológico que apenas reconocemos pero a las que todavía tenemos que describir. Primero debemos exponer el punto metafísico, y luego por fin, nos podremos ocupar del punto psicológico.

Considérese lo que deberíamos decir acerca de las condiciones de identidad de la creencia citada en

37) Tom cree que el espía de menor estatura es una mujer.

Si tratamos esta atribución relacionalmente, como debemos hacerlo si queremos saber si la creencia de Tom es verdadera o falsa, entonces la preposición que él cree, dado el mundo en el que está inserto, es acerca de (en el sentido débil) alguna persona real, supongamos que Rosa Klebb, de manera que su creencia es verdadera. Si alguna otra persona, el Diminuto Traidor, el espía de menor estatura que estaba entonces en *ese* mundo, Tom habría tenido una creencia falsa. ¿Diríamos, no obstante, que en *ese* mundo Tom hubiera tenido la misma creencia, o una distinta? Aquí tenemos que tener cuidado en distinguir el estado psicológico restringido o actitud nocional de la actitud proposicional. También tenemos que tener cuidado en distinguir las reflexiones acerca de cúal *habría podido ser* el caso de lo que le *puede venir bien* al ejemplo. Porque observe que el estado psicológico limitado de Tom podría mantenerse constante en los aspectos pertinentes por, digamos, un año, durante el cual el "título" de espía de menor estatura cambiaría de

mano, sin el conocimiento de Tom, por supuesto. Durante meses Rosa Klebb fue la espía de menor estatura, hasta que Tom el Diminuto aceptó el rublo de la KGB. Durante esos primeros meses, Tom creyó algo falso. Deben ser proposiciones distintas. En realidad tantas proposiciones diferentes como usted quiera, durante ese año, dependiendo de cuán finas corte las tajadas del engañoso regalo. Los cambios en Tom, sin embargo, a medida de que cada proposición evanescente pasa como un relámpago, como una serie de proposiciones reales (el 1° de enero el espía de menor estatura es una mujer, etc.) va seguida, de repente, por una serie de proposiciones falsas, que son lo que Geach llamaría cambios Cambridge, tal como ese cambio que le acontece a uno cuando deja repentinamente de tener la propiedad de estar más cerca del Polo Norte que el más anciano pionero viviente nacido en Utah. No todas las actitudes nocionales están relacionadas así con las actitudes proposicionales, por supuesto.

38) Tom cree que el miembro más joven de la clase 1950 de Harvard se graduó con honores.

Cuál proposición cree en este caso no cambia de un día para el otro; más aun, acerca de quienquiera que sea esta creencia (en el sentido débil), es acerca de esa persona para toda la eternidad. No podemos decir que otra persona podría *llegar a ser* en el futuro el objeto (en el sentido débil) de esa creencia, pero *quizá* podamos encontrarle sentido a la sugerencia de que, contrariamente a los hechos, alguna otra persona *podría haber sido* el objeto de *esa misma creencia*. Para ver qué implica esto, repasemos las posibilidades. Determine la actitud *nocional* de Tom. Primeramente, hay mundos posibles en los que Tom, con su actitud nocional o estado psicológico limitado, no *cree absolutamente ninguna proposición*; éstos son mundos por ejemplo sin Harvard, y también sin ninguna Harvard Gemela. Segundo, hay mundos posibles en los que Tom cree una proposición no acerca de Harvard sino acerca de alguna Harvard Gemela y su graduado más joven de 1950 Gemelo. Tercero, hay mundos en los que Tom cree la mismísima proposición que cree en el mundo real, pero en el cual otra persona es el graduado más joven. La identidad de la proposición está unida al atributo; más joven graduación, y no a su portador. Y en 37) las proposiciones creídas están unidas a menor estatura-espionaje, no a los portadores de título. Esa es, en todo caso, la doctrina que se debe sostener si 37) y 38) han de ser diferenciados por el punto de vista de la Referencia Especial de las creencias auténticamente específicas. Esta doctrina nos exige encontrar el sentido de los reclamos de que es posible que algún otro haya podido ser el espía de menor estatura (el 1° de enero, el 2 de enero...), y es posible que algún otro haya sido el graduado más joven de la promoción 1950 de Harvard. Considere ahora

39) Tom cree que la persona que dejó caer la envoltura de la goma de mascar (quienquiera que pueda ser) es desprolija y descuidada.

Supongamos (puesto que algunos dicen que hay diferencia) que Tom no vio a esa persona sino la envolutra descartada de la goma de mascar. Ahora bien, da la casualidad que la creencia de Tom es, en este caso, acerca (en el sentido débil) de cierto individuo, quienquiera que haya arrojado la envoltura de la goma de mascar, y la creencia será acerca de ese individuo por toda

la eternidad; pero si la historia hubiera sido sólo un tanto diferente, si algún otro hubiera dejado caer la envoltura de la goma de mascar, entonces la mismísima creencia de Tom (la mismísima actitud nocional y la mismísima actitud proposicional) hubiera podido ser acerca de otra persona. ¿Podría haber sido acerca de otro si Tom hubiera visto la envoltura cuando se la dejaba caer? ¿Por qué no? ¿Cuál sería la diferencia si Tom veía a la persona? Si Tom veía a la persona, sin duda

40) Tom cree que la persona a quien vio arrojar la envoltura de la goma de mascar (quienquiera que fuese) es un desprolijo descuidado,

también será verdad. Esta creencia es débilmente acerca del mismo individuo descuidado, pero si la historia hubiera sido un poco diferente, también habría sido acerca de algún otro. Es en este punto que el creyente en el punto de vista de la Referencia Especial de la creencia *de re* interviene para insistir en que, por el contrario, si la historia hubiera sido un poco distinta, si Tom hubiera visto a algún otro dejar caer la envoltura de la goma de mascar, *habría tenido una creencia diferente.* Habría sido una creencia diferente porque hubiera sido acerca de (en el sentido *fuerte*) una persona diferente. Es fácil confundir aquí dos afirmaciones diferentes. Es tentador suponer que lo que el teórico *de re* piensa es esto: si la historia hubiera sido un poco diferente, si una persona alta y delgada hubiera sido vista arrojando la envoltura de la goma de mascar, en lugar de una persona baja y gorda, las percepciones de Tom hubieran sido muy diferentes de manera que él habría estado en un estado psicológico limitado muy diferente. Pero aunque esto sería normalmente cierto, esto no es lo que el teórico *de re* piensa; el teórico *de re* no está haciendo una afirmacin acerca del estado limitado de Tom sino acerca de la actitud proposicional de éste: si la historia hubiera sido un poco diferente, si el hermano gemelo de la persona baja y gorda hubiera dejado caer la envoltura de la goma de mascar en circunstancias no muy distinguibles, la actitud nocional de Tom podría haber sido exactamente la misma, suponemos, pero habría tenido una creencia diferente, una actitud proposicional distinta por completo, porque —sólo porque— su estado nocional fue causado directamente (de alguna manera especial) por un individuo distinto.[34]

La metafísica de este punto de vista se puede enfocar claramente por medio de un ejemplo imaginario. Supongamos que Tom hubiera estado lle-

[34] No todos los creyentes en la teoría causal analizan los ejemplos de la misma manera. Vendler (1981) insiste en que aun en el caso en que no veo dejar caer la envoltura de la goma de mascar, *puesto que solamente una persona podría haber dejado caer* esa envoltura —puesto que sólo una persona podría haber hecho esa kripkeana "inserción en la historia"— mi creencia es rígida y fuertemente *acerca de* esa persona. ¿No estamos familiarizados en un sentido muy real con el esclavo por otra parte desconocido que dejó la huella de su pie en la tumba del Rey Tut? ¿O con el escriba que talló *ese* jeroglífico especial en *esa* piedra hace 4000 años?" (pág. 73). Sospecho que la posición aparentemente extrema de Vendler es la única posición estable que un teórico causal puede adoptar. Pero quizá yo haya entendido mal a Vendler: quizás *alguna otra persona* podría haber dejado caer *esa* envoltura de goma de mascar, pero *nadie más* podría haber dejado *esa* huella del pie. Entonces, la creencia de Tom acerca de quien dejó la pisada en el asfalto húmedo en el sentido de que él es un desprolijo descuidado es directamente acerca de *ese* individuo de manera que su creencia acerca de quien dejó caer la envoltura no es directamente acerca de *él*. No por nada, varias personas ingeniosas llamaron a la teoría causal la doctrina del *pecado [Sinn]* original.

vando consigo durante años una moneda de la suerte. Tom no tiene ningún nombre para esa moneda, pero podemos llamarla Amy. Tom llevó a Amy consigo a España, la guarda en su mesita de noche cuando duerme, y demás. Una noche, mientras Tom duerme, una persona perversa saca a Amy de la mesita de luz y reemplaza a Amy con una impostora, Beth. Al día siguiente Tom acaricia amorosamente a Beth, la pone en su bolsillo y se va a trabajar.

41) Tom cree que una vez llevó la moneda a España en el bolsillo.

Esta oración verdadera acerca de Tom afirma que él tiene una creencia falsa determinada acerca de Beth, pero si la historia hubiera sido un poco diferente, si la persona perversa no hubiera intervenido, esta creencia misma habría sido una creencia cierta acerca de Amy. Pero según la teoría de la Referencia Especial de las creencias *de re*, hay otras creencias por considerar:

42) Tom cree de Amy que la tiene en el bolsillo.

43) Tom cree de Beth que una vez la llevó a España.

Estas son creencias fuertemente acerca de Amy, la primera falsa, la segunda cierta, y ser acerca de Amy es *esencial* para ellas. Acerca de ellas no podemos decir: si la historia hubiera sido un poco de otro modo *ellas* habrían sido acerca de Beth. Si así fuera, entonces hay que considerar:

44) Tom cree de Beth que la tiene en el bolsillo.

45) Tom cree de Beth que la llevó una vez a España.

Estas son creencias fuertemente acerca de Beth, la primera verdadera, la segunda falsa y el que sean acerca de Beth es *esencial* para ellas. Acerca de ellas no podemos decir: si la historia hubiera sido un poco de otro modo *ellas* habrían sido acerca de Amy. ¿Por qué tendríamos que adoptar este punto de vista? No puede ser porque necesitamos admitir que Tom tiene creencias acerca de ambas monedas, puesto que en (41) su creencia es (débilmente) acerca de Beth y

46) Tom cree que tiene en el bolsillo la moneda que llevó a España

le atribuye a Tom (débilmente) una creencia acerca de Amy. Concedamos que en todos los casos de creencia que son débilmente acerca de esos objetos en virtud de ser esos objetos los únicos que satisfacen una descripción, podemos introducir, para marcar esto, la frase "quienquiera que sea". A menudo esto sonará muy extraño, como

47) Tom cree que su esposa (quienquiera que sea) es una nadadora excelente

pero sólo porque la implicación pragmática de la frase introducida sugiere una posibilidad muy exótica, a menos que se la tome meramente para sugerir que *el que habla* en (47) no puede reconocer a la esposa de Tom entre un grupo de personas.

¿Hay entonces otra razón por la cual deberíamos adoptar este punto de vista? No puede ser porque necesitemos distinguir una clase especial de creencias *de re* fuertemente acerca de objetos para señalar una clase diferenciada de casos donde *cuantificar*, puesto que podemos cuantificar allí donde una creencia es sólo débilmente acerca de un individuo, aunque esa práctica sea a menudo sumamente engañosa. Sería muy engañoso, por ejemplo, decir, señalando a Rosa Klebb durante su reino como el espía de menor estatura: "Tom cree que ella es una mujer". Sería también muy engañoso, no obstante, que la persona perversa mostrara a Amy y dijera "Tom una vez llevó

esta moneda a España". Estas afirmaciones serían engañosas puesto que la implicación pragmática normal de una afirmación relacional de esta clase es que el creyente puede identificar o volver a identificar el objeto (o propiedad) en cuestión. Pero nada acerca de la adquisición de una creencia podría *garantizar* esto contra todos los contratiempos futuros de manera que *cualquier* teoría de la "creencia *de re*" autorizaría afirmaciones relacionales potencialmente engañosas (véase Schiffer, 1978, págs. 179, 188).

Si todavía parece que hay una *variedad* evasiva *de creencia* con la que no se ha sido justo, esto se debe probablemente a un tentador diagnóstico equivocado de las clases de ejemplos que se encuentran en la literatura. Extraer unos pocos ejemplos más puede ser útil para conjurar el fantasma de la creencia *de re* (en este sentido imaginario que lo distingue como una subvariedad) de una vez por todas. Supongamos que estoy participando de una reunión de comisión y se me ocurre que la persona más joven que está en la habitación (quienquiera que sea: media docena de personas presentes son candidatas posibles) nació después de la muerte de Franklin D. Roosevelt.

Llamemos a ese pensamiento mío Pensamiento A. Ahora bien, en el sentido débil de "acerca" el Pensamiento A es sobre una de las personas presentes, pero no sé cuál.

Las miro una por una y me pregunto, por ej., "Ese de allí, Bill: ¿es posible que el Pensamiento A sea *acerca de él*? Llamemos a *este* pensamiento mío Pensamiento B. Ahora bien, *con seguridad* (uno siente) que el Pensamiento B es *acerca de Bill* en un sentido mucho más directo, íntimo, fuerte que el Pensamiento A, aun si el Penamiento A resulta ser realmente acerca de Bill. Creo que ésta es una ilusión. Hay sólo una diferencia de grado entre el Pensamiento A y el Pensamiento B y su relación con Bill. El Pensamiento B es (débilmente) acerca de quienquiera que sea la única persona a la que estoy mirando y cuyo nombre creo que es Bill y.... por todo el tiempo que usted quiera. Sin duda, Bill es el único que responde a esa descripción, pero si su hermano gemelo hubiera tomado su lugar sin yo saberlo, el Pensamiento B no hubiera sido acerca de Bill sino acerca de su hermano. O más probablemente, en esa eventualidad, yo estaría en un estado mental como el del pobre Tom en la Pizzería de Shakey, de manera que no se dispone de ninguna interpretación psicológicamente inteligible de mis actitudes *proposicionales*.

Otro ejemplo con distinto sabor. George es el asesino de Smith como Gracie lo sabe bien y entra corriendo para contarle que Hoover piensa que él lo hizo. Alarmado, George quiere saber cómo Hoover pensó en él. Las razones de Gracie son: Hoover sabe que la única persona que le tiró a Smith con una 38 dejó tres impresiones digitales no registrables en la ventana y está en estos momentos en posesión del dinero de la billetera de Smith es el asesino de éste, puesto que George es el único que responde a la descripción de Hoover, Hoover cree de George que él lo hizo. "No, Gracie", dice George, "Hoover sólo sabe que *quienquiera* que corresponda a esta descripción es el asesino de Smith. No sabe que *yo* respondo a la descripción, de manera que no sabe que yo soy el asesino de Smith. Despiértenme cuando sepan que se sospecha de *mí*" (véase Sosa, 1970). George ha hecho un mal diagnóstico de esta situación, sin embargo, ya que considera el caso en el que Poirot reúne a todos los invitados a la fiesta en el salón y dice "todavía no sé *quién es el asesino*; ni siquiera

tengo un sopechoso, pero he inferido que el asesino, quienquiera que sea, es la única persona en la sala que tiene consigo una copia de la llave de la despensa". A continuación hay búsqueda, identificación y arresto. No es cierto que George esté a salvo mientras las creencias de Hoover tienen la forma de *quienquiera que responda a la descripción B es el asesino de Smith*, puesto que si la descripción D es algo parecido a "la única persona en el bar de Clansy con barro amarillo en los zapatos" enseguida empezará el lío.[35] Uno es un sospechoso (mínimo) si satisface cualquier descripción definida. Hoover se pone a dar con el asesino de Smith. Se deduce trivialmente que el asesino de Smith es un sospechoso mínimo (porque responde a la descripción "asesino de Smith") aun en la situación en que Hoover se ve totalmente desconcertado, pero cree meramente que el crimen fue cometido por un solo acusado. Esta sería una consecuencia objetable sólo si hubiera alguna manera por principios de distinguir los sospechosos mínimos de los auténticos o los sospechosos verdaderos o *de re*, pero no la hay. De este modo, como lo sugiere Quine, la distinción psicológica aparentemente aguda entre

48) Hoover cree que alguien asesinó a Smith

y

49) Hay alguien acerca de quien Hoover cree que asesinó a Smith

se derrumba (persiste la diferencia lógica en el compromiso ontológico de quien habla). Sigue siendo cierto que en el caso en el que Hoover está confundido le negaría naturalmente a la prensa que hubiera alguien acerca de quien el creía que fuera el asesino. Lo que estaría negando realmente es que sabe más que lo que sabe cualquiera, que sabe sólo que se ha cometido el crimen. Por cierto que no está negando que tiene una creencia *de re* directamente acerca de cierto individuo en el sentido de que él es el asesino, una creencia que adquirió por medio de alguna afinidad cognitiva íntima con ese individuo, pues supongamos que Hoover luchó con el asesino en la escena del crimen, a plena luz del día pero no tiene idea de quién era la persona con la que luchó; con seguridad, en la teoría causal de la creencia *de re* de cualquiera, esa persona es la que él cree que es el asesino, pero sería muy solapado de su parte afirmar que tiene un sospechoso (véase Sosa 1970, págs. 894 y sigs.).

La unión de la identidad de creencia y la referencia de creencia con una condición causal es una propuesta plausible porque en la mayor parte de los casos los distintos grados de intimidad causal en el pasado se pueden utilizar

[35] Consideremos el contraste entre

a) Entro en una cabina telefónica y encuentro una moneda en el recipiente que devuelve las monedas. Creo que quienquiera que usó la cabina por última vez dejó una moneda en el recipiente.

b) Entro en una cabina telefónica, hago una llamada y dejo deliberadamente una moneda en el receptáculo. Creo que quienquiera que use la cabina después encontrará una moneda en el recipiente que devuelve las monedas.

¿Mi creencia en (b) es *ya* acerca de alguna persona en especial? ¡Pero no tengo la menor idea de quién es o será! (Véase Harman, 1977). ¿Y qué? (Véase Searle, 1979). No tengo la menor idea acerca de quién es mi creencia en a) tampoco y en realidad es mucho más probable que yo descubra el objeto de mi creencia en b) que en a). Si creo que quienquiera que dejó la moneda secuestró a Jones, el secuestrador está probablemente a salvo del todo. Si creo que quienquiera que encuentre la moneda es el secuestrador (es una señal en el esquema de intercambio de rescate) la captura es más probable.

para distinguir las relaciones más débiles de las más fuertes entre los creyentes y sus objetos, de manera tal que las relaciones más fuertes tienen implicaciones en la conducta futura, pero en las situaciones desusadas las implicaciones normales no se mantienen. A partir del hecho de que podemos producir dislocaciones, vemos que la exigencia causal no es necesaria o suficiente por sí misma; los efectos son tan importantes como las causas. Lo que se necesita es la creación de un objeto nocional en el mundo nocional del sujeto. Esto no sucederá habitualmente en ausencia del intercambio causal. Es improbable que algo "se ponga en la posición" necesaria como para que alguien tenga creencias acerca de ese algo sin haber estado en interacción causal de *algún* tipo con el creyente, pero en casos especiales podemos provocar este resultado (el perchero de Westwood Village, la puerta, el acusado de Poirot). Los casos especiales atraen la atención hacia una independencia en teoría —si bien pocas veces en la práctica— entre los estados psicológicamente salientes y sus credenciales metafísicas.

El creyente en la creencia *de re* tiene que decidir si el concepto en discusión ha de jugar o no un papel notable en las explicaciones conductistas (véanse, por ejemplo, Morton, 1975; Burge, 1977). Un punto de vista de la creencia *de re* no supondría para nada que algo acerca de la conducta probable de Tom surge de la verdad de

50) Tom cree del hombre que le está dando la mano que es un asesino múltiple fugitivo fuertemente armado.

Este punto de vista admite lo que podría ser llamada la opacidad psicológica de la transparencia semántica (podemos no saber en absoluto acerca de qué son nuestras creencias), y si bien no veo ningún obstáculo para definir esa variedad de actitud proposicional, no le veo ninguna utilidad a semejante concepto puesto que nada interesante parecería resultar de una atribución verdadera de semejante creencia. Supongamos, según este punto de vista, que algo, *a* es tenido por mí por ser *F*. Esto no significa que yo no crea también que *a* es *no-F*; y si también creo que *a* es *G* ello no significa que yo crea que *a* es *F* y *G*; ni siquiera surge del hecho que yo creo que *a* es *la única F*. La premisa de la búsqueda de la creencia *de re* era que había relaciones muy interesantes e importantes entre los creyentes y los objetos de sus creencias —relaciones que teníamos motivos para captar en nuestras teorías— pero este final de la búsqueda nos conecta con relaciones de poco interés. Si eso es así, ya no veo ningún motivo para negar que uno cree *acerca del* espía de menor estatura que él es un espía. Se dispone de los instrumentos formales para formular esas afirmaciones relacionales, y una vez que se ha afirmado que dos objetos están relacionados, por más mínimamente que sea, como creyente y objeto, se puede pasar a afirmar cualquier otra cosa acerca del estado mental del creyente o la situación del objeto que pueda ser pertinente al hecho de tramar y explicar las futuras carreras pertinentes de ambos.

Si, al rehuir este punto de vista, se busca un punto de vista de la creencia *de re* como una especie de fenómeno psicológicamente diferente, no puede ser entonces una teoría bien llamada, puesto que tendrá que ser una teoría de distinciones dentro de la psicología de la actitud *nocional*. Si *a* es un objeto *de mi mundo nocional* que yo creo ser *F*, resulta en realidad que yo no creo también que *a* sea *no-F*, y las otras implicaciones citadas antes también

ocupan su lugar, pero sólo porque los objetos nocionales son "lo creado" por las creencias de alguien (véase Scheffer, 1978, pág. 180). Una vez que hemos creado esos seres, podemos ver con qué cosas reales se alinean (si es que lo hacen con alguna),pero no desde ninguna posición de acceso privilegiada en nuestros propios casos. Existe una intuición muy poderosa de que puede ser de ambas maneras: que podemos definir una especie de *acerquidad* que es *tanto* una verdadera relación entre un creyente y algo en el mundo *y* algo a lo que el acceso del creyente es perfecto. Evans llama a esto el principio de Russell: *no es posible formular un juicio acerca de un objeto sin saber acerca de qué objeto se está formulando un juicio.* En el caso de Russell el intento por preservar esta intuición frente a los tipos de dificultades aquí encontradas lo llevó a su doctrina de la sabiduría por conocimiento y, por tanto, inevitablemente a su punto de vista de que podríamos sólo formular un juicio *acerca de* ciertos objetos abstractos o ciertos estados internos propios. El Principio es:

> Cada vez que ocurre una relación de suposición o juicio, los términos con los que la mente que supone o juzga está relacionada mediante la relación de suponer o juzgar, deben ser términos con los que la mente en cuestión está familiarizada. (Russell 1959, pág. 221).

Aquí el término "término" llena exactamente el vacío y prepara el camino para precisamente la clase de teoría que Chisholm aborrece en la cita que abre esta sección: una teoría que supone que "la mente no puede ir más allá del círculo de sus propias ideas". La salida es abandonar el Principio de Russell (véase Burge, 1979) y con él la idea de una clase especial de creencia *de re* (y otras actitudes) que sean íntima y fuertemente *acerca de* sus objetos. Sin embargo todavía queda algo de verdad en la idea de Russell acerca de una relación especial entre un creyente y algunas *de las cosas con las que piensa*, pero para discutir este punto hay que volverse hacia la psicología de la actitud nocional y más particularmente hacia la cuestión "operativa" de cómo diseñar un ser cognitivo con el tipo de mundo nocional que comúnmente tenemos. En ese terreno, protegido ahora de alguna de las doctrinas engañosas acerca del fenómeno mítico de las creencias *de re* y *de dicto*, surgen diferencias fenomenológicas que atraviesan los límites tentativos de la reciente literatura filosófica y contienen cierta promesa de producir una visión más clara de la organización cognitiva.

Puesto que este trabajo ya ha superado varias veces la extensión originariamente propuesta reservaré un examen detallado de estas diferencias para otra ocasión y sólo daré una lista de los puntos que juzgo más promisorios.

Distintas maneras de pensar en algo. Este es un conjunto puramente nocional de diferencias: ninguna de las maneras implica la existencia de algo en lo que uno está pensando. Vendler (1976, 1984) tiene algunas reflexiones valiosas acerca de este tema (conjuntamente con su no propuesto *reductio ad absurdum* de la teoría causal de la referencia para la creencia).

La diferencia entre el pensar y el creer (episódicos). En cualquier momento tenemos *creencias* acerca de muchas cosas sobre las que no podemos *pensar*; no porque las creencias sean inconscientes o tácitas o impensables, sino simplemente porque estamos temporalmente incapacitados para *acceder*

a aquello a lo que podemos desear acceder. ¿Se puede *pensar* en la persona que nos enseñó la división tradicional? Si se "busca a esa persona en la memoria" se descubrirá sin duda una reserva secreta de creencias acerca de ella. Cualquier teoría psicológica plausible de la acción debe tener una explicación de cómo reconocemos las cosas y de cómo nos mantenemos al tanto de ellas (véase Morton 1975), y esto exige una teoría de las estrategias y procesos que usamos para explotar nuestras creencias en nuestro pensamiento.

La diferencia entre la representación explícita y la virtual (véase capítulo 6). Cuando subo a un coche alquilado y parto, espero que esté en perfectas condiciones de funcionamiento y por tanto espero que el neumático delantero derecho tenga llantas seguras. Me sorprendería descubrir lo contrario. No sólo no he pensado *conscientemente* acerca del neumático delantero derecho sino que casi con seguridad, no he representado inconscientemente (pero sí explícitamente) la llanta del neumático delantero derecho como un ítem aparte de creencia. La diferencia que establece *haber prestado atención a algo* (real o meramente nocional) no es la diferencia entre *de re* y *de dicto*; es una diferencia verdadera de alguna importancia en la psicología.

La diferencia entre las creencias lingüísticamente contaminadas y el resto: lo que llamo opiniones y creencias en "How to Change Your Mind" (en *Brainstorms*). Ningún perro podría tener la opinión de que era viernes o de que el perro más chico era un perro. Algunos han supuesto que esto significa que los animales que no tienen lenguaje son incapaces de la creencia *de dicto*. Es importante reconocer la independencia de esta diferencia, de los temas en juego en los debates *de re/dedicto*.

La diferencia entre los objetos artefactual o nítidamente nocionales y otros objetos nocionales. Podemos conjurar algo imaginario sólo para fantasear o para resolver un problema, por ej., para diseñar una casa de ensueño o imaginar qué clase de coche comprar. No tenemos siempre la intención de que los mundos nocionales que construimos dentro de nuestros mundos nocionales sean *ficción*. Por ejemplo, al encontrar el cadáver de Smith podríamos reconstruir el crimen en nuestra imaginación: una manera diferente de pensar en el asesino de Smith.

Confío en que esté claro que pasar por alto estas diferencias ha contribuido a la confusión en las discusiones del *de re* y del *de dicto*. Considérese simplemente el caso vergonzoso de creer que el espía de menor estatura es un espía. Se supone comúnmente que todos sabemos a qué estado mental se alude en este ejemplo, pero en realidad hay muchas posibilidades diferentes que no aparecen en la literatura. ¿Cree Tom que el espía de menor estatura es un espía en virtud de nada más que una cuota normal de lógica o tiene que tener también cierto tipo de ociosidad y alguna propensión para reflexionar acerca de las tautologías? (¿Tendríamos que decir que todos creemos esto y también que el árbol más alto es un árbol y así *ad infinitum*?). ¿Cree Tom que la ballena más pequeña es una ballena? O para cambiar el rumbo, ¿cuál es la relación entre creer que ese hombre es un espía y pensar que ese hombre es un espía? A menudo se invoca el rechazo sincero como un signo revelador de la incredulidad: ¿qué es lo que se percibe mejor que lo que se expone? Y así sucesivamente. Cuando se disponga de buenas explicaciones psicológicas de estos fenómenos de la actitud nocional se resolverán algunos de los enig-

mas acerca de la referencia y otros quedarán desautorizados. dudo de que quede algún residuo.[36]

Reflexiones: Acerca de la acerquidad

Fodor comienza su libro pionero *Psychological Explanation* (1968a) con un poco de auto-burla:

> Pienso que muchos filósofos sostienen secretamente el punto de vista de que hay algo profundamente (es decir, conceptualmente) erróneo en la psicología, pero que un filósofo que tenga cierta preparación en las técnicas del análisis lingüístico y una tarde libre lo podría corregir.
>
> Hace unos años me encontré con una tarde libre (pág. vii).

En 1978 Stich y yo, que estábamos participando en un taller sobre la filosofía de la psicología de la Fulbright de Woodfield en la Universidad de California, tuvimos una corazonada parecida: era evidente para nosotros que había algo profundamente (es decir, conceptualmente) equivocado en el edificio total de la teoría del deseo/creencia *de re/de dicto* en la filosofía y pensamos que podríamos pasar algunas semanas haciendo juntos el trabajo sucio de corregirlo todo. Durante esas pocas semanes descubrimos que a pesar de nuestro desánimo compartido, teníamos diferencias residuales e insolubles acerca de cómo tratar los temas. Después de más de un año de trabajo, presentamos trabajos separados para la antología 1981 de Woodfield: "Más allá de la creencia" yo y Stich *On the Ascription of Content*, que años más tarde se convirtió en el centro de su punto de vista en *From Folk Psychology to Cognitive Science: The Case Against Belief* (1983).

A pesar del año de trabajo, "Más allá de la creencia" es, sin duda, un proyecto inconcluso, y hasta algunos de sus admiradores se han sentido inseguros acerca de cuáles son sus mensajes. De manera que haré una lista de las que considero son sus afirmaciones principales, las discutiré una por una y luego diré algo más acerca de dónde considero que esto deja las cosas.

1) *Las proposiciones:* en este momento no hay ningún punto de vista admitido y estable acerca de las proposiciones o las actitudes proposicionales en el que se pueda confiar. Las dos escuelas de pensamiento principales, la de las proposiciones como cosas parecidas a las creaciones y la de las proposiciones como conjuntos de palabras posibles, son sumamente incompatibles y apelan a intuiciones muy diferentes.

[36] Agradezco a todas las almas pacientes que han tratado de ayudarme a encontrar la salida a este proyecto. Además de todos los mencionados en las otras notas y bibliografía quiero reconocer la ayuda de Peter Alexander, David Hirschmann, Christopher Peacocke, Pat Hayes, John Haugeland, Robert Moore, Zenon Pylyshyn, Paul Benacerraf, y Dagfinn Follesdal. Esta investigación fue apoyada por una Beca N.E.H. y por la National Science Foundation (BNS 78-24671) y la Alfred P. Sloan Foundation.

2) *Los mundos nocionales:* el muy transitado camino que va de la "psicología de la actitud proposicional" a la "psicología del lenguaje del pensamiento" es una ruta al mundo de la fantasía. Una psicología más realista no usa "proposiciones" (ni sus semejantes más cercanos) para caracterizar directamente los mecanismos filosóficos; más bien los utiliza para caracterizarlos en forma indirecta: caracterizando de manera directa el "mundo" con el cual esos mecanismos están diseñados para tratar.

3) *El principio de Russell:* la suposición común de que *no es posible formular un juicio acerca de un objeto sin saber acerca de qué objeto se está formulando el juicio* —llamado el Principio de Russell por Evans— debe ser dejada de lado. Debido a su atracción intuitiva e introspectiva es la fuente de la parte del león de los infortunios de los teóricos.

4) *Los de re/de dicto*: no hay ningún punto de vista estable o coherente de la así llamada diferencia *de re/de dicto*. Las explicaciones que apelan a esa supuesta distinción están viciadas de una ambiguedad evasiva entre una distinción psicológicamente notable y proyectable y una distinción metafísicamente precisa pero psicológicamente inerte.

Estas tesis predominantemente negativas estaban destinadas a perturbar la complacencia de aquellos filósofos que pensaron que los enigmas (acerca de Ortcutt, del espía de menor estatura, del Planeta Tierra Gemelo) se podían manejar dentro de la tradición presumiblemente ortodoxa, digamos russelliana. En mi opinión esta ortodoxia es una ilusión, un subproducto de la sociología de la disciplina: cuando los expertos escriben para expertos, tienden a errar por el lado de presuponer un acuerdo mayor del que realmente existe acerca de los hechos básicos. La literatura filosófica acerca de las proposiciones ha sido desde el principio el trabajo de los filósofos del lenguaje y de la lógica, principalmente, cuyas preocupaciones han estado relacionadas con la filosofía en forma tangencial. Para ellos (hasta hace poco tiempo) los enigmas acerca de las actitudes proposicionales eran obstáculos aparentemente periféricos en una estructura por otra parte muy linda, y se ocupaban de ellos inventando sólo la suficiente psicología plausible como para taparlos. Podría haber resultado, pero no resultó: la filosofía de la psicología impulsada por los intereses de la filosofía del lenguaje no encaja bien. De manera que ahora tenemos a los teóricos, cuyas intuiciones han sido mancilladas por una tradición provisoria pero autoritaria, que han estado rivalizando los unos con los otros durante una generación, presuponiendo erróneamente que había un entendimiento común acerca de lo que se suponía que hace el concepto central de una proposición.

Las proposiciones

Existe todavía una creencia extendida en la ortodoxia imaginaria estable, pero la mía no es en absoluto la única idea iconoclasta; la mezcla de contención e invención en la literatura (que ha crecido rápidamente desde que terminé "Más allá de la creencia" en 1980) la ha puesto fuera del alcance de todos menos de los especialistas atrevidos, lo que es probablemente lo mismo. A los demás se les alienta para que miren hacia otro lado hasta que armemos nuestro acto.

Y, sin embargo, estos otros podrían proporcionarle una infusión de realidad importante a una subdisciplina cada vez más innata y artesanal. Después de todo, si hay algo que los demás académicos parecen necesitar de los filósofos, es una teoría de las proposiciones o por lo menos una teoría de algo parecido a las proposiciones. El término "información" se usa a menudo como un sustantivo de materia, en inglés, como si la información fuera un tipo de sustancia que se pudiera mover, almacenar, comprimir, picar. Los teóricos de disciplinas tan variadas como la neurociencia, la etología, la economía y la crítica literaria fingen saber cómo medir y caracterizar los montones de este material que llamamos información. Según los medios masivos de comunicación, vivimos ahora en la Era de la Información, pero la realidad es que no tenemos ninguna comprensión sólida y mutuamente reconocida acerca de qué es la información y en qué clase de paquetes habría que medirla.

La información caracterizada por la teoría formal de la información, medida en bits o bytes (1 byte = 8 bits), subyace bajo toda transferencia y procesamiento de la información, y esta es por cierto una mercancía bien entendida que puede ser partida y luego almacenada o trasladada en porciones bien diferenciadas. Pero ese no es en absoluto el concepto al que debemos recurrir cuando nos preocupamos por la filtración de información (importante) en el gobierno, el coste de obtener información en la toma de decisiones administrativas, el torrente de información en el que nosotros y nuestros proyectos parecen estar ahogándose (Dennett, 1986b), o cuando hablamos de modelos de procesamiento de información del sistema nervioso o en la psicología cognitiva (Dennett, 1969, págs. 185-89). La información medida en bits es neutra en contenido. He aquí un recordatorio gráfico de esto: un solo videodisco puede almacenar los suficientes gigabytes de información como para grabar toda la Enciclopedia Grolier —o tres horas de Bugs Bunny. Hay aproximadamente la misma *cantidad de información* en cada uno. cuando queremos saber *qué información* (sobre qué temas, con qué contenido) se transfirió cuando Philby habló con los rusos, o cuando el ojo de la rana le habló al cerebro de ésta, necesitamos un concepto diferente. Tenemos un nombre para él —*información semántica*— pero a pesar de las tentativas ingeniosas (si bien fracasadas) de desarrollar el concepto indispensable como una extensión del concepto teórico de información (Dretske, 1981, Sayre, 1986, Dennett, 1986a), todavía no tenemos una manera mejor de diferenciar las porciones del maravilloso material que hablar de una maenra u otra de las proposiciones.

Las proposiciones, como el medio final de la transferencia de información, como trocitos de realidad (o ficción) de lenguaje neutro, lógicamente diferentes, independientes para la observación, continúan jugando un rol fundamental como los elementos atómicos en muchas investigaciones teóricas y proyectos prácticos. Como las *pes* y las *qs* del cálculo proposicional y sus herederas, las proposiciones siempre tienen a mano su atomicidad no analizada. Los objetos estructurados del cálculo del predicado y sus herederos y relaciones presuponen un lenguaje canónico —si bien imaginario— y tienden, por tanto, a otorgarle una confiabilidad espuria sobre cierta forma de oracionismo. Mientras tanto, los psicólogos cognitivos y sociales se dedican a catalogar las *creencias* de sus objetos anotando sus respuestas a cuestionarios;

y en la inteligencia artificial, el uso del cálculo del predicado para captar las creencias putativas del sujeto que se usa como modelo es una tradición bien arraigada sin rivales serios, claros. (Para un panorama crítico de los problemas de las proposiciones en la inteligencia artificial, véase Dennett "Cognitive Wheels, the Frame Problem of *AI*" (1984c) y "The Logical Geography of Computational Approaches: A View from the East Pole" (1986a). Si a la teoría filosófica de las proposiciones le va tan mal como lo afirmo, ¿corren peligro todos estos proyectos o están todos los enigmas filosíficos sumergidos *tranquilamente a salvo bajo una ola de cláusulas "para todo uso"*?

Mi defensa de la actitud intencional revela mi propia confianza en las proposiciones y en las atribuciones de la actitud proposicional y demuestra al mismo tiempo los límites del peso que creo se les puede poner encima. Las proposiciones se portarían mucho mejor si no tratáramos de sujetarlas tanto a la "psicología de la actitud proposicional". Es decir, que el origen principal de los problemas con las proposiciones es el caudal de intuición acerca de la comprensión de ellas. Una vez que se afloja el control de la comprensión tratando a las proposiciones como si "midieran" los estados psicológicos (como lo recomienda Churchland) sólo indirecta y aproximadamente, juegan bastante bien su papel limitado. Mientras se reconozca que el debate acerca de las proposiciones no es nada más que una *capa heurística* (Dennet, 1969, pág. 80), una aproximación útil —si bien a veces traicionera— que es sistemáticamente incapaz de dar un resultado exacto, se puede eludir el problema con lo que es en esencia la pretensión de que toda la información puede modelarse sobre el *relato*, mandando un *dictum* de *A* a *B*, Lo que el ojo de la rana le dice al cerebro de ésta no es nada que se pueda captar íntegramente en una oración, y si algún espía vende un juego de copias secretas en Ginebra, la información así trasmitida puede ser exacta e igualmente incapaz de "ser explicitada" en ninguna fórmula del cálculo del predicado. [37] Pero comúnmente se puede encontrar una oración que destile de manera útil el *quid* (no la esencia en ningún sentido más fuerte) de la información a la que uno se refiere. Se puede afirmar que esta oración expresa, en una primera aproximación, la proposición en cuestión.

¿Qué es la proposición misma y cómo se la puede identificar exactamente? El camino de la proposición como oración canónica está bloqueado por el hecho de que cualquier información es relativa a un sistema intencional determinado —la rana, el cerebro de la rana, la organización de la defensa nacional, el espía maestro— y en virtud de los diferentes predicamentos de esos sistemas intencionales su capacidad de captación puede diferir en maneras inconmensurables que desafían cualquier *lingua franca* de transmisión. El camino alternativo, la proposición como un conjunto de mundos posibles, tiene la gran virtud de reconocer y basarse precisamente en esta relatividad del agente. [38]

[37] Véase Dennett, 1983b, para un mayor desarrollo de la analogía entre la ciencia cognitiva y el contraespionaje.

[38] Stalnaker (1984) es quien hace una exploración más perspicaz y sistemática de la perspectiva. La objeción estándar y presumiblemente abrumadora al tratamiento de las proposiciones creídas de los mundos posibles es que, ya que una posibilidad es lo que es, puede haber (para cual-

Tratar la proposición como nada más que una clasificación particular exhaustiva de posibilidades perceptibles del agente, permite en principio una determinación más exacta y realista de la "proposición real creída por el agente", pero dada su inconmensurabilidad, esa precisión así obtenida limita a la generalidad de las afirmaciones interagentes que se pueden expresar. Lo que Tom, Pierre, Sherlock y Boris tienen en común (ver capítulo 3) no se puede expresar exactamente en términos del mundo posible, puesto que inevitablemente separarán sus conjuntos de mundos posibles por la vía de conceptos idiosincráticos algo diferentes de *francés*, *asesino* y *plaza Trafalgar*. [39] Sostengo que no hay nada *exactamente* en común entre Tom, Pierre, Sherlock y Boris.

Si esto es así, hay que volver a plantear ligeramente la búsqueda de una teoría estable de las proposiciones. Cualquier fraccionamiento útil de la información semántica fijará un estándar algo arbitrario y potencialmene distorsionante, de manera que debemos complicar la atractiva propuesta de Churchland de que tratemos a las proposiciones del mismo modo en que los físicos tratan a los números. Las proposiciones como maneras de "medir" la información semántica por medio de la plenitud del tema, *resultan ser más parecidas a los dólares que a los números*. Tal como "¿cuánto vale eso en dólares estadounidenses?" formula una pregunta unificadoramente útil a pesar de las frecuentes ocasiones en las que la respuesta distorsiona la realidad en la que estamos interesados, así: "¿qué proposición (en el Esquema Estándard *P*) almacena / transmite / expresa eso?" podría explotar un campo de experimentación valioso, algo sistemático, si bien a menudo inflexible. Sólo los norteamericanos ingenuos confunden la primera pregunta con "¿cuánto vale eso en dinero *verdadero*?" y sería igualmente ingenuo considerar un estándar de determinación de la proposición, por bien establecido que esté, como siquiera una aproximación al modo en que la información semántica se divide *realmente*. *No hay unidades reales, naturales, universales ya sea de valor económico o de información semántica.*

Presumiblemente no nos desalienta ni nos asombra el descubrimiento de que no hay unidades naturales, universales de valor económico. ¿Cuánto vale una cabra viva? Pues bien, ¿dónde y cuándo? En la Francia rural actual tiene mucho valor pero ¿es más o menos que su valor corriente en China o el valor de una camioneta Chevy de segunda mano en Louisiana en 1972? ¿Cuánto vale hoy la misma cabra viva entregada en la oficina privada de un

---

quier creyente único) sólo una proposición lógicamente verdadera o necesaria —verdadera en *todos* los mundos posibles— mientras que los matemáticos, por ejemplo, parecerían creer, y dudar, por cierto, de muchas verdades lógicas diferentes. Stalnaker argumenta que este fenómeno se puede manejar distinguiendo las distintas creencias de los matemáticos acerca de *determinadas fórmulas* en el sentido de que son "expresiones" de la verdad lógica solitaria. Mientras esta táctica les parece a algunos una estratagema desesperada, yo le encuentro una motivación independiente. Las distinciones de grano fino entre las verdades lógicas son accesibles sólo para (son comprensibles para, tienen importancia para) los sistemas intencionales que utilizan lenguajes. Por tanto, los estados psicológicos en discusión son esos estados contaminados verbalmente a los que llamo *opiniones*, y el ascenso semántico de Stalnaker es una manera promisoria de lograrlo.

[39] Curiosamente, se puede extraer la misma moraleja de la consideración de otro francés de ficción en Londres: el Pierre de Kripke, quien podría creer o no que Londres es linda (Kripke, 1979. Entre las muchas réplicas a Kripke, véase especialmente Marcus, 1983).

corredor de Wall Street? Se puede hacer una evaluación del "valor" de la cabra (digamos en dólares, en 1968) en distintos momentos y lugares, para distintas personas, trabajando con los tipos de cambio, la tasa de inflación, los precios de mercado, etcétera. (¿Quién ganó más en un año promedio: uno de los generales de César o un juez de la Corte Suprema?) Doy por sentado que nadie supone que semejantes ejercicios se refieren a valores económicos reales y únicos. ¿Por qué se debería seguir pensando que hay contenidos de los estados psicológicos únicos y reales, unidades de información semántica singularmente bien articuladas?

Los mundos nocionales

La psicología del mundo nocional, al intentar captar la contribución orgánica a la creencia aislada, es un enfoque de lo que ha venido a ser llamado "semántica del rol conceptual limitado (Field, 1977, 1978; Loar, 1981; Fodor, 1987. Para críticas esclarecedoras del enfoque véase Harman, de próxima aparición a; Stalnaker, inédito). Queremos hacerle justicia a la intuición de que la información de la que estamos hablando —la información semántica— está *en* el organismo de una manera que importa (véase, por ejemplo, Stich, 1983). Esto se puede dramatizar por medio de la exigencia de que la información se pueda mandar de *A* a *B* mandando al organismo (aislado) de *A* a *B* y contando con que quienes la reciben hagan el mejor trabajo posible de interpretación radical. Esto lo harán reconstruyendo el mundo nocional del organismo. Hasta el punto en que el resultado no está determinado —es decir, hasta el punto de que los mundos nocionales rivales estén igualmente bien sustentados por todos los datos acerca del organismo disponible en B— la información en cuestión simplemente *no está allí.*

Algunos (véase especialmente Stalnaker, inédito) han dudado de que este método de reconstrucción fuera tan fecundo como lo he afirmado. ¿Podrían los intérpretes marcianos, al recibir un curpo humano en estado comatoso por correo interplanetario, reconstruir su mundo nocional tan bien como yo lo afirmo? Obsérvese que el trabajo de los marcianos no es oficialmente más difícil que —por cierto sólo un caso especial de— el trabajo imaginario conocido de la traducción quineana radical [o la "interpretación radical" de Davidson (1973) o la de Lewis (1974)]. Lo que hace especial el caso es que el cuerpo está efectivamente separado de cualquier influencia ambiental (sea en Marte o en la Tierra), pero *ex hypothesi* los marcianos pueden determinar cómo el cuerpo, una vez despierto, *respondería* a cualquier estímulo que cualquier entorno podría proveer, de manera de poder llevar a cabo, hipotéticamente, tantos experimentos *gavagai*, como deseen, y cosechar los beneficios. Presumiblemente utilizarían sus cálculos para inducir al cuerpo, hipotéticamente, a relatar en su idioma nativo tanto de su biografía como pueda reunir. (Hofstadter describe el ejercicio en detalle en *A Conversation with Einstein's Brain"* en Hofstadter y Dennett, 1981.) Suponiendo que los marcianos tomaran las medidas disponibles para controlar las mentiras y los olvidos, su examen debería producir, como yo lo sostengo, una descripción de un mundo nocional que no distingue a Boston de la Boston Gemela pero que es por otra parte notablemente específico.

Una vez más no estoy proponiendo la psicología del mundo nocional como una metodología a ser seguida diligentemente por los psicólogos empíricos, sino sólo como una manera de ser explícito, cuando ocurren los pocos casos de dislocación, acerca de los compromisos y presunciones de la teoría. La oblicuidad de las descripciones de los mundos posibles tiene su utilidad, entonces para caracterizar al pueblo entre quienes creen en Papá Noel, sin comprometer al teórico con un Papá Noel en el mundo ni con un "Papá Noel" en cada cerebro. Y el formato del mundo nocional nos recuerda que, como Jackendoff (1985) lo ha expresado, "la información está en la mente del observador".

El principio de Russell

El papel pernicioso del principio de Russell me ha quedado gradualmente mejor enfocado, gracias en buena medida al ataque sostenido contra él (bajo el rótulo más amplio de "racionalismo con sentido") por parte de Millikan (1984, 1986). En "Más allá de la creencia", por ejemplo, me dediqué a describir un Fodor poco objetivo cuando yo me lo imaginaba afirmando "encontrar un texto en lenguaje mental *dado* en el hardware" y por tanto pensé que yo era meramente hipotético cuando decía que "tendríamos tanta razón en dudar de la existencia de lo dado en este caso como en el caso de la fenomenología". Pero como lo demuestra el capítulo 8, yo estaba más cerca de la verdad de lo que suponía al poner esta afirmación en boca de Fodor, y Fodor está muy lejos de estar solo al querer preservar el acceso privilegiado de la primera persona para alguna *comprensión* fregeana de las proposiciones.

El *de re* / *de dicto*

Las tentativas de definir esta diferencia en los términos del linaje causal fracasan en los procedimientos instructivos. Podríamos volver a bautizar el concepto de acerquidad fuerte (*de re*) como una *referencia original directa* para ayudarnos a ver que es uno de una familia de conceptos perennemente atractivos, incluyendo la *intencionalidad original* y *la funcionalidad original* (véase el cap. 8) todas las cuales fracasan ante la posibilidad de lo que llamo hechos históricos inertes. Los orígenes son importantes, pero también lo son las propensiones o disposiciones corrientes, y si se está interesado en las propiedades provechosamente proyectables de las cosas, se deben tratar todas las condiciones que parecen un retroceso como indicios sólo normalmente muy confiables —síntomas— más que como criterios. Como admití en "Más allá de la creencia" es muy posible definir una versión de la creencia *de re* en términos causales; es igualmente fácil, como veremos en el capítulo 8, definir una versión de la *intencionalidad original* de manera tal que algunas cosas tengan *de verdad* estados con (ciertos) contenidos, y otras cosas de conducta idéntica no los tengan, debido a la ilegitimidad de sus orígenes. En ninguno de los casos se capta nada de valor por medio del test, a menos que se le asigne un valor algo místico al dato genealógico bruto. La manera estándar de defi-

nir ese misticismo es atacar a la oposición por su verificacionismo ("lo que nadie podría *verificar* sería empero *verdad*") pero esto desvía inadecuadamente mi crítica. Estoy completamente preparado para reconocer la existencia de verdades inútiles, que no se pueden verificar —las llamo verdades inertes, a pesar de todo— pero insisto en que no pueden tener la importancia teórica que sus defensores les confieren; lo que no se puede verificar ya no puede "ser importante" y nos engañamos cuando les damos mucha trascendencia a diferencias teóricas en cuya importancia no se puede confiar.

"Más allá de la creencia" termina con una mirada algo ansiosa a algunos trabajos inconclusos de la filosofía de la psicología que adoptan un aspecto diferente una vez que uno se quita de encima el estado hipnótico de la tradición; y me parece que varios trabajos nuevos proporcionan letreros indicadores a seguir al menos acerca de cómo avanzar sobre estas perplejidades residuales. Además del trabajo filosófico de Millikan y Stalnaker ya citado, recomiendo muy favorablemente dos libros por dos lingüistas psicológica y filosóficamente sagaces: Jackendorf (1983) y Fauçonnier (1985).

6

Los estilos de representación mental*

Hace más de treinta años en *The Concept of Mind* (1949), Gilbert Ryle atacó una visión de la mente que llamó el mito intelectualista. Esta era la idea de que las mentes están compuestas principalmente por episodios tales como pensar pensamientos privados, consultar preceptos y recetas, la aplicación de verdades generales a circunstancias particulares y la deducción posterior de las implicaciones acerca de esas particularidades. Este punto de vista acerca de la mente no era menos absurdo por ser tradicional, según la opinión de Ryle, y su ataque fue una combinación vigorosa —si bien no rigurosa— de ridículo y *reductio*. Por supuesto, sabía perfectamente bien que en realidad no existen fenómenos tales como la memorización de normas, la consulta posterior de esas normas, la deducción consciente de las conclusiones a partir de las premisas, y otras cosas por el estilo, pero él pensó que considerar estas actividades escolares como modelos de toda conducta inteligente era retroceder. En realidad, argumentaba, la existencia misma de tales prácticas humanas públicas como la consideración consciente y deliberada a partir de conjuntos de premisas escritas en pizarras *depende* de que los agentes en cuestión tengan conjuntos de talentos mentales cabales. Si se intentara explicar estas competencias anteriores y más fundamentales como basadas a su vez en todavía otro proceso intelectual, esta vez un proceso de cálculo *interno* que implicara buscar proposiciones, sacar conclusiones de ellas y cosas por el estilo, se estaría dando el primer paso de una regresión infinita.

Era un ataque poderoso y retóricamente atractivo, pero pocos años después fue rechazado lisa y llanamente por aquellos filósofos y otros teóricos que vieron la esperanza de una psicología cognitiva o más ampliamente de una "ciencia cognitiva", una teoría de la mente que estaba muy cerca en espíritu del punto de vista que Ryle ridiculizaba. En realidad, la ideología reinante de la ciencia cognitiva se planta de manera tan desafiante contra Ryle que se la podría llamar con justicia *ciencia intelectualista*. Parece ser exactamente la clase de cosa que Ryle proclamaba como obstinada. La ciencia cognitiva habla abierta y descaradamente de representaciones mentales internas y de cálculos y de otras operaciones realizadas en estas representaciones internas. Proclama que la mente es un aparato de computación y co-

* "Los estilos de la representación mental aparecieron por primera vez en *Proceedings of The Aristotelian Society*, vol. LXXXIII 1982 / 1983. Reimpreso por cortesía del comp. de la Aristotelian Society.

mo lo ha expresado Jerry Fodor, un ideólogo rector de la ciencia cognitiva, "no hay computación sin representación" (Fodor, 1975).

¿Cómo hizo este movimiento nuevo para dejar de lado tan animosamente el ataque de Ryle? Parte de la respuesta es que lo que Ryle atacaba no era un punto de vista sino una mezcolanza de varios puntos de vista. Los científicos cognitivos sofisticados pueden esquivar muchos de sus dardos tan fácilmente que tal vez su respeto por el resto de su arsenal haya disminuido excesivamente. Ryle bailó completamente a su gusto sobre el cadáver del dualismo cartesiano, por ejemplo, pero la ciencia cognitiva es abiertamente materialista o fiscalista, de una manera sofisticada que Ryle aparentemente subestimó y quizá ni siquiera consideró. De manera que la ciencia cognitiva no se preocupa para nada por los "fantasmas de la máquina". Sus hipótesis son francamente mecánicas y no como Ryle lo querría "paramecánicas", misteriosamente seudomecánicas. La ciencia cognitiva no le guarda ninguna fidelidad al "acceso privilegiado", otro de los cucos de Ryle. Por cierto, se supone que la mayor parte de las representaciones mentales de las que habla son completamente innacesibles para la conciencia del agente. En su mayor parte es una doctrina de representaciones mentales *inconscientes*. De manera que éste no es el intelectualismo del teatro cartesiano interno en el que todo sucede en el escenario de la conciencia. Este es un intelectualismo de "bambalinas", y según el punto de vista de estos nuevos teóricos tanto mejor para él.

Pero había mucho más que eso en el ataque de Ryle. El sospechaba profundamente —y creo que por una buena razón— de cualquier afirmación hecha en favor de las representaciones internas, ya fuera que se supusiera que estaban encarnadas materialmente, mecánicamente manipuladas, fuera de la conciencia o no, porque pensaba que esas postulaciones de las representaciones internas eran incoherentes en todas sus formas.

Que esta característica del punto de vista de Ryle no ha pasado por cierto inadvertida. ¿Por qué entonces ha sido pasada por alto tan ampliamente? Con seguridad, la contribución principal a la convicción de que Ryle debe estar equivocado acerca de este punto es la creciente influencia de la metáfora del ordenador en ese terreno. El eslogan dice que la mente es como un ordenador y un ordenador es por cierto una manipuladora sin inteligencia de las representaciones explícitas internas. Allí no hay regresiones infinitas, con seguridad, puesto que allí se sientan los ordenadores. Todo lo que sea real es posible, de manera que los manipuladores de las representaciones internas no son las máquinas de movimiento perpetuo que Ryle quisiera que pensáramos que son. Los científicos cognitivos se han sentido cómodos entonces al hablar de los sistemas y subsistemas de procesamiento de la información que utilizan tipos de representación interna. Estas representaciones son como palabras y oraciones, como mapas y cuadros. Son lo bastante parecidas a estas representaciones conocidas como para que sea apropiado llamarlas representaciones, pero también son lo bastante distintas de las palabras y los cuadros, así sucesivamente como para evadir la máquina de regresión infinita de Ryle. Esa es, en todo caso, la comprensión común no que ya haya sido defendida correctamente.

Propongo estudiar con mayor atención el tema, puesto que mientras que creo que la metáfora del ordenador, adecuadamente usada, *puede* liberar al teórico de las preocupaciones de Ryle, a menudo abusan de ella los ideólogos de la ciencia cognitiva. Admitamos que es evidente que los ordenadores *de algún modo* representan cosas utilizando representaciones internas de esas cosas. De manera que es *posible* (de algún modo) que los cerebros representen cosas por medio de representaciones internas de ellas. Pero si hemos de trasladarnos más allá de esa pequeña parte del espacio metafísico y entender realmente *cómo* el cerebro podría representar por analogía con la representación del ordenador, sería mejor que tuviéramos claro cómo representan los ordenadores. Hay varias maneras o estilos de la representación por ordenador y sólo algunos de ellos son modelos plausibles para las maneras o estilos de representación mental en el cerebro.

Cuando los científicos cognitivos hablaron acerca de las representaciones se comprometieron comúnmente con determinado punto de vista acerca de la sintaxis, para usar el término taquigráfico. Es decir, que supusieron que hablaban de las representaciones, acerca de las cuales se pueden hacer las siguientes clases de distinciones: hay elementos estructurales que son símbolos; hay *signos* múltiples de *tipos* de representación (donde estos tipos se diferencian sintáctica y no semánticamente); hay reglas de formación o reglas de composición —algo parecido a una gramática— de manera que se puedan formar representaciones grandes a partir de las pequeñas; y el significado de las representaciones más grandes es una función de los significados de sus partes.

Este pesado compromiso con una imagen sintáctica, se toma a menudo de muy buena gana, pero si se lo acepta es únicamente sobre la base del razonamiento apriorístico, puesto que hasta donde puedo alcanzar a ver, mientras que ha habido mucha especulación interesante, todavía casi no hay ninguna evidencia empírica que tienda a confirmar ninguna hipótesis sustantiva acerca de la naturaleza de esta supuesta sintaxis de la representación mental (véase Stabler, 1983). Esto no significa en absoluto negar que se han obtenido pruebas excelentes y significativas por medio de la investigación cognitiva, sino más bien afirmar que hasta la fecha estas pruebas lo han sido sólo a nivel semántico. Es decir, que han sido pruebas acerca de aquella información en la que en cierto modo confían distintos procesos cognitivos y no la evidencia acerca de cómo se produce esta confianza (Dennett, 1983d).

Por ejemplo todavía no podemos decir si distintos datos que fueron implicados de una manera u otra en distintas actividades y competencias cognitivas, están representados "explícitamente" o "implícitamente" en el sistema cognitivo humano. Una razón por la cual no podemos decir esto es la confusión acerca de cómo usar los términos "explícito" e "implícito". La gente que trabaja en esto ha querido expresar cosas completamente diferentes por medio de estos términos y aquí trataré de aclarar algo la situación ofreciendo algunas definiciones, no de cómo se usan los términos, sino de cómo deberían usarse algunos términos.

Digamos que la información está representada *explícitamente* en un sistema si y sólo si existe realmente en el lugar funcionalmente pertinente del

sistema un objeto físicamente estructurado, una *fórmula* o *sarta* o *simbología* de algunos miembros de un sistema (o "lenguaje") de elementos para los que hay una semántica o interpretación y una disposición (un mecanismo de algún tipo) para leer o analizar la fórmula. Esta definición de la representación explícita es exigente, pero aún deja lugar a una gran variedad de sistemas de representación. No es necesario que sean sistemas lineales, secuenciales ni como oraciones, sino que podrían por ejemplo ser "sistemas lectores de mapas" o "intérpretes de diagramas".

Asumamos entonces que para que la información esté representada *implícitamente*, pretenderemos decir que está *implicada* lógicamente por algo que está almacenado explícitamente. Ahora bien: *implícitamente* así definido no significa lo que se podría entender qué significa: "potencialmente explícito". Todos los teoremas de Euclides están implícitos en los axiomas y definiciones. Y si se tiene una máquina Euclides mecánica —si se tienen los axiomas y definiciones almacenados explícitamente en la máquina y ésta puede producir teoremas profusamente— entonces los teoremas que produce y, por tanto vuelve explícitos, estaban implícitos todo el tiempo en el sistema, eran implicados por los axiomas. Pero también lo estaban muchísimos otros teoremas que la máquina puede *no ser* capaz de producir. Están todos implícitos en el sistema, dada su representación explícita de los axiomas, pero sólo un subconjunto apropiado de ellos son potencialmente explícitos. (Esto no se refiere a la "calidad de incompleto": pienso en las limitaciones más mundanas de las piezas medianas de hardware que obedecen la velocidad einsteiniana límite durante lapsos de vida relativamente breves.)

Es interesante la pregunta de si el concepto de representaciones *potencialmente explícitas* es de mayor utilidad para la ciencia cognitiva que el concepto de las representaciones (meramente) implícitas. Expresado de otra manera: ¿puede algún ítem de información que esté *meramente* implícito en algún sistema citado alguna vez (con cualquier propósito explicativo) estar en una explicación cognitiva o intencional de cualquier suceso? Creo que hay una corriente oculta de convicciones fuerte pero tácita en el sentido de que sólo si se lo vuelve explícito, sólo si es realmente *generado* por algo parecido a la máquina de Euclides puede un ítem de información *jugar un rol*. Aparentemente, la idea es que para dar algún resultado, para tener algún peso, por así decirlo, una información debe tener cierto peso, debe tener una encarnadura física y ¿cuál podría ser ésta sino su representación o expresión explícita? Sospecho, por el contrario, que esto es casi al revés. Las representaciones explícitas, solas (consideradas aisladamente de los sistemas que las pueden utilizar) pueden ser trocitos del universo admirablemente notables, de donde despedir fotones o moléculas neurotransmisoras o canicas, pero son por sí solas completamente inertes como portadoras de información en el sentido que necesitamos. Se *convierten* en portadoras de información sólo cuando se les asignan papeles en sistemas más grandes, cuando aquellas de sus características en virtud de las cuales las llamamos explícitas juegan papeles problemáticos en el mejor de los casos.

Se podría decir muy bien que la representación implícita no es en absoluto una representación; sólo la representación explícita, lo es. Pero es enton-

ces cuando se debe pasar a observar que si esto es lo que entendemos por "representación", hay modos de retener o hasta enviar información en un sistema que no implican representarlo. Después de todo, un espía puede mandar un mensaje de *A* a *B* indirectamente, enviando premisas explícitas de las que se deduce el mensaje propuesto, la información-a-enviar.[1] Otro punto importante que hay que recordar acerca del almacenaje implícito de información es que no tiene ningún límite superior. No le hace falta ocupar más espacio para almacenar más información implícita.

De manera que *implícito* depende de *explícito*. Pero en el sentido de "tácito" lo que voy a utilizar es lo contrario: *explícito* depende de *tácito*. Esto es lo que Ryle quería decir cuando proclamaba que probar cosas explícitamente (en pizarras, etcétera) dependía de que un agente tuviera mucho *"know how"*, que no podría ser explicado por sí mismo en términos de la representación explícita en el agente de ninguna norma o receta, puesto que para poder manipular *esas* reglas y recetas tendría que haber un agente interno con el *"know how"* necesario como para manejar esos términos explícitos, y eso llevaría a un retroceso infinito. Ryle vio que en el fondo tiene que haber un sistema que tenga *meramente* el *"know how"*. Si se puede decir que en algo *representa* a su *"know how"*, no lo debe representar ni explícita ni implícitamente —en el sentido recién definido— sino tácitamente. El *"know how"* tiene que estar incorporado al sistema en alguna forma que no le exija estar representando (explícitamente) en él. La gente usa muchas veces la palabra "implícito" para describir la posesión de información; lo que quieren decir es lo que yo quiero decir con "tácito".

Ryle pensaba que el retroceso de los representadores tenía que detenerse en alguna parte, con los sistemas que tienen solamente un *"know how"* tácito. Tenía razón. Pero también pensaba que era evidente que personas enteras no estaban compuestas de subsistemas más pequeños que prepresentan explícitamente algo por sí mismos, y estaba equivocado al respecto. No es *evidente*, como lo demuestran varias décadas de psicología cognitiva. Sin embargo, podría ser cierto —o en todo caso estar más cerca de la verdad de lo que la ideología común supone.

Todos estos términos —"explícito", "potencialmente explícito", "implícito" y "tácito"— han de ser diferenciados de "consciente" e "inconsciente". De este modo, lo que uno se representa conscientemente para uno mismo es, en el mejor de los casos, una prueba indirecta de lo que podría estar representado explícitamente en uno de manera inconsciente. Hasta donde le concierne a la ciencia cognitiva, los fenómenos importantes son las representaciones mentales inconscientes explícitas. De modo que cuando Chomsky (1980a) habla acerca de la representación explícita de la gramática de alguien en su cabeza, no se refiere ciertamente a la representación consciente de esa gramática. Se presume que es inconsciente y totalmente inaccesible para el sujeto. Pero también quiero decir que no está meramente tácita en el funcionamiento o capacidad del sistema, ni está tampoco meramente implícita

[1] Esta posibilidad es ingeniosamente explotada por Jorge Luis Borges en su clásico cuento "El jardín de los senderos que se bifurcan" en *Labyrinths: Selected Stories and Other Writings*, compilado por Donald A. Yates y James E. Irby, Nueva York, Nuevas Directivas, 1982.

en algo "más básico" que esté explícito en la cabeza. El se propone adoptar la *línea dura*: la gramática está representada por sí misma inconsciente pero explícitamente en la cabeza (véase Chomsky 1980a, 1980b).

Podemos entender la línea dura comparándola con un caso paradigmático de observancia explícita *consciente* de las reglas. Tenga en cuenta a los jugadores de bridge y su relación con una conocida regla práctica de este juego.

¡El tercero juega alto!

Aquí, en esta misma página, hay una representación explícita de la regla. Está representada con frecuencia explícitamente (simplemente en estas tres palabras en inglés) en libros y artículos sobre bridge. También se pueden escuchar sus símbolos gritados de un lado a otro de las mesas de bridge. La regla le impone al tercero de los cuatro jugadores jugar su carta más alta para ganar cualquier baza en el palo que está en juego (ésta es una regla táctica, no una regla del juego. Se puede saber jugar al bridge y no conocer la regla; se puede jugar *buen* bridge y no "conocer" la regla).

Tenga en cuenta el caso más extremo. Esta es la persona que conscientemente y hasta tímida y explícitamente consulta la regla, quien, cuando le llega el turno de jugar una carta, piensa para sí (¡tal vez hasta mueve los labios!): "Pues bien, veamos, creo que hay una regla para esto. Sí, ¡el tercero juega alto! '¿Soy el tercero? Uno, dos, tres; sí, soy el tercero. Se supone que debo jugar la carta más alta del palo que está en juego. Esa sería mi *J*", y entonces juega su *J*. Ese sería el caso de seguir explícitamente una regla: extraer la regla de la memoria, presentarla en el "espacio de trabajo" de la conciencia, examinarla, estudiarla, controlarla para ver si reúne las condiciones, y al ver que sí lo hace, "disparar" la actividad. Ahora bien, la afirmación de Ryle era que cualquiera que piense que éste es un buen modelo para el mentamiento humano de todas clases, desde atarse los cordones de los zapatos hasta comprender una oración en nuestra lengua activa, está —digamos— sumido en la oscuridad. Pero esto es lo que piensan los científicos cognitivos, por lo menos algunos de ellos. Podemos ver cuál es su tema comparando nuestro primer jugador de bridge con algunos otros tipos conocidos.

Tome en cuenta luego al jugador "intuitivo" de bridge. Supongamos que nuestro jugador intuitivo nunca en su vida escuchó la regla, y las palabras de la regla nunca se le ocurrieron en ningún idioma que él sepa. Nunca se "refleja" en ella, de manera que cuando piensa qué carta jugar ciertamente no sigue la regla de manera consciente o explícita. Eso le deja sólo la posición en la que la mayoría de nosotros estamos con respecto a las reglas de nuestro idioma nativo. No debe deducirse que no está siguiendo *in*conscientemente alguna versión explícita de la regla. Supongamos en cualquier caso que sus disposiciones para la conducta en el juego de cartas (la conducta para *elegir* las cartas— hemos de pasar por alto los entrecejos fruncidos, las demoras y el mascullado *sotto voce* que la acompañan) no se pueden distinguir de las del primer jugador de bridge. La hipótesis de la línea dura es que el procesamiento del relato de bambalinas de elección "intuitiva" de cartas de este jugador se parece mucho al relato "introspectivo" que nuestro primer jugador podría hacer.

Finalmente, vean el tercer caso: un jugador que combina las características de los dos primeros y, por lo tanto es mucho mejor jugador de bridge que el primero y tal vez mejor que el segundo también. Esta persona conoce la regla, pero es lo suficientemente lista como para darse cuenta de que esa regla tiene excepciones y que no hay que seguirla ciegamente. Puede pensar en las razones que la fundamentan y acerca de si ésta es una buena oportunidad para aplicarla.

El tercer jugador sería un modelo mucho mejor pero para ser adoptado por la psicología cognitiva (y sospecho que Ryle tenía en mente esta especie de consultor de reglas cuando dejó de lado el proyecto) porque carece de una propiedad importante que el primer jugador tenía; la estupidez. Una característica sistemáticamente importante revelada por el seguimiento de la regla de nuestro primer jugador es la posibilidad de almacenar y "actuar según" algo sin entenderlo en realidad. Es el peor tipo de actividad escolar, la memorización mecánica, la que proporciona el mejor modelo para la ciencia cognitiva, porque tiene la refinada característica de desconectar la memoria de la comprensión. Esta clase de memoria es simplemente almacenamiento bruto (como un cantante que memorice la letra de una canción rusa sin tener la menor idea de lo que quiere decir).

Parece que esto es todo lo que necesitamos si estamos tratando de explicar la comprensión en términos de almacenaje y manipulación. Queremos que nuestros almacenadores y manipuladores sean más estúpidos que nuestro entendedor (del que son partes propias); de otro modo entraríamos en una represión ryleana. Los almacenadores y manipuladores deben tener ciertamente algún "*know-how*"; hasta nuestro primer jugador de bridge sabe lo bastante como para saber cómo aplicar la regla, y esto no es ninguna nadería; sólo piénsese en los jugadores de bridge que no parecen poder meterse esta sencilla regla en sus duras cabezas. A su vez este "*know-how*" podría ser meramente tácito *o* basarse en algún proceso ulterior interno de seguimiento de reglas, de horizontes aun más estrechos y mayor estupidez (véase *Brainstorms*, capítulos 5 y 7). En principio, la posibilidad de hacer concluir este retroceso en un número limitado de pasos es una comprensión rectora fundamental de la ciencia cognitiva. ¿Pero se usará de verdad el retroceso? Al final debe concluir con el "*know-how*" meramente tácito pero podría resultar que Ryle tuviera razón después de todo acerca de la "magnitud" de los más numerosos conocedores meramente tácitos: todas las personas? ¿Podría ser virtualmente todo el "*know-how*" de *bambalinas* meramente tácito en la organización del sistema? ¿Cuán poderoso puede ser un sistema de "representación" tácito?

Esta es una pregunta difícil de contestar, en parte debido a que al término crítico o "tácito" sólo se le ha dado hasta ahora una definición impresionista, ostentosa. No hemos comprendido bien, en realidad, lo que debería significar. Téngase en cuenta una pregunta clave: ¿representa una calculadora de bolsillo las "verdades de la aritmética" explícita, implícita o tácitamente? Una diminuta calculadora de mano *nos da acceso* a una virtual infinidad de hechos aritméticos, pero ¿en qué sentido están éstos "almacenados" en ella? Si se observa bien el hardware, no se encuentran proposiciones numé-

ricas escritas en código en su interior. La única representación explícita evidente de los números es que estén impresos en el botón de entrada de datos, o durante la salida, desplegados en letras de cristal líquidas en el visor.

¿Pero hay con seguridad una representación explícita ulterior oculta para el usuario? Téngase en cuenta lo que ocurre cuando se le da el problema a la calculadora: $6 \times 7 = ?$ Supongamos que la calculadora efectúa la multiplicación sumando velozmente (en notación binaria) $7 + 7 + 7 + 7 + 7 + 7$ y que durante el proceso real retiene en su acumulador o "buffer" los totales provisionales de cada suma sucesiva. De esta manera podemos distinguir claramente el proceso por el que atraviesa el multiplicar 7×6. En el primer caso los resultados provisionales son 14, 21, 28, 35, mientras que en el segundo son 12, 18, 24, 30, 36. Con seguridad que ésta es una representación de *números* explícita y sistemática pero, ¿dónde representa la calculadora cualquier *proposición* aritmética verdadera? Su mecanismo interno está dispuesto de manera tal que tiene la fantástica propiedad dispositiva de *contestar correctamente las preguntas aritméticas*. Lo hace sin siquiera *consultar* ningún dato aritmético o reglas de funcionamiento almacenadas en ella. Por supuesto que fue diseñada por ingenieros que conocían las verdades de la aritmética y las reglas del cálculo aritmético y que se ocuparon de que el artefacto funcionara como para "hacer honor" a todas esas verdades y reglas. De manera que la calculadora es un artefacto con la competencia dispositiva como para producir respuestas explícitas a preguntas explícitas (de un modo tal que estas verdades están *potencialmente explícitas* en ella), pero realiza esto sin apoyarse en ninguna representación explícita dentro de ella, excepto las representaciones de las preguntas y respuestas que aparecen en sus ejes de entrada y salida y una diversidad de resultados provisionales. Las verdades de la aritmética potencialmente explícitas en ella no están así *implícitas*, puesto que no hay verdades explícitas en ella de las cuales éstas sean las implicaciones.

Por ridículo que pueda parecer al principio, vale la pena comparar este punto de vista de la calculadora de bolsillo con el punto de vista de Ryle acerca de los seres humanos. Ciertamente que Ryle es el enemigo de la representación interna, pero tiene el buen sentido de reconocer lo que podría llamarse la representación explícita *periférica*, en los ejes de entrada y salida de datos de los seres humanos (!) como lo ejemplifican las conocidas categorías ryleanas como los ensayos *sotto voce*, hablar solo y sin mover los labios, recordarse a uno mismo los resultados provisionales antes de seguir con la tarea que se tiene entre manos. El punto de vista de Ryle, tal y como lo entiendo, es que exactamente como no hay ninguna representación más profunda, oculta pero explícita, de nada en la calculadora de bolsillo, en virtud de la cual podamos decir y pensar las cosas que hacemos.

Una característica interesante del proceso de diseño que produce cosas como las calculadoras de mano como productos, es que los diseñadores *empiezan* normalmente con una especificación perfectamente explícita de las verdades a respetar y las reglas a seguir. Eventualmente tienen éxito en crear un artefacto que "obedezca" las reglas y "respete" las verdades sin que él mismo las represente en absoluto explícitamente. ¿Tiene este proceso algo análogo a él en el desarrollo mental humano? Ocasionalmente, los seres humanos adquieren habilidades que están gobernadas al principio (o por lo me-

nos así nos lo asegura enfáticamente la "introspección") por la consulta completamente explícita de reglas ensayadas explícitamente —tal como ocurría con nuestro primer jugador de bridge— pero con la práctica estas habilidades, *en cierto modo* se "automatizan": la jugadora de tenis ya no se murmura instrucciones a ella misma mientras se prepara para un golpe de revés, la persona que acaba de adquirir "fluidez" para hablar un segundo idioma ya no verifica "conscientemente" para hablar un segundo idioma, ya no verifica "conscientemente" para asegurarse que el adjetivo concuerde en género con el sustantivo que modifica.[2] ¿Qué está pasando en estos fenómenos probablemente afines? Una posibilidad, muy vívidamente esbozada por Fodor (1968b) es que con esa automatización se trata meramente de *ocultar* el seguimiento explícito de la receta que está debajo del acceso normalmente consciente. (Su ejemplo es atarse los zapatos; se supone que sumergida debajo del acceso consciente hay una receta explícita llamada "Cómo atarse los zapatos". Esta es rescatada por un subsistema de lectura y seguimiento de la receta igualmente oculto, que gobierna el verdadero proceso de atarse los zapatos tal como ocurría con nuestro primer jugador de bridge.) Otra posibilidad, sugerida por el proceso de diseño de la calculadora, es que la práctica es de algún modo un proceso análogo de autodiseño parcial: produce como su análogo, un "artefacto" que obedece las reglas sin consultar ninguna de sus expresiones.

Esta última posibilidad puede parecer probable sólo para sistemas cuyas tareas y, por tanto cuyas reglas de funcionamiento, son tan estáticas e invariables como aquellas "fuertemente conectadas" a una calculadora. Pero es completamente posible diseñar sistemas que puedan cambiar el modo, pasando de "seguir" un conjunto de reglas tácitas a "seguir" otro. Por ejemplo, se puede hacer que un ascensor automático siga un conjunto de reglas de las nueve a las diecisiete los días hábiles y un conjunto de reglas diferentes durante las horas que no son punta y las del fin de semana. Y por supuesto que se puede hacer lo que haga sin que ninguno de los dos conjuntos de reglas esté explícitamente interpretado en él; lo único que necesita es un interruptor controlado con un reloj para llevarlo de un lado a otro entre dos sistemas de control diferentes, cada uno de los cuales representa tácitamente un conjunto de reglas de funcionamiento. Esto nos da un ejemplo simple de lo que podemos llamar *representación tácita transitoria*. Todas las reglas están representadas tácitamente todo el tiempo, pero según el estado del sistema, sólo un conjunto de reglas está tácitamente representado *como siendo seguido* en algún momento. O se podría afirmar igualmente que todo el sistema representa tácita y permanentemente la regla "Siga el sistema de la regla *R* los días hábiles de 9 a 17 y el sistema de la regla *R* en otros momentos", y el estado del sistema en cualquier momento representa transitoria y tácitamente el momento de la semana, lo que es una manera de proporcionar un vehículo (pero no un vehículo de representación *explícita*) para proposiciones indicativas, tales como "Ahora estamos en el fin de semana".

[2] Uno de los casos más fascinantes de este cambio hacia la automaticidad es el autoaprendizaje de un prodigio en cálculos del que informa Hunter, 1962.

Podemos imaginarnos sistemas similares en los animales. Podemos imaginarnos, por ejemplo, un animal que sea tanto acuático como terrestre y que cuando está en tierra obedece un conjunto de reglas y que cuando está en el agua obedece otro. Simplemente el hecho de *mojarse* podría ser el disparador para cambiar el estado interno de un conjunto de reglas al otro.

Hasta ahora los sistemas que he descrito cambian de un estado a otro por medio de un simple interruptor, un sencillo tecleo del "transductor" en alguna característica del entorno (o el mero paso del tiempo si el transductor es un reloj), pero podríamos tener un mecanismo de cambio más complicado, de manera que el sistema de reglas que estuviera transitoria y tácitamente representado dependiera de rasgos distales complejos del entorno. Si el complicado mecanismo de análisis perceptivo conduce el sistema a sus distintos estados, entonces no hay límite aparente para la especificidad o la complejidad del estado en el mundo, por ejemplo, que pudiera ser representado tácitamente por el estado actual de ese sistema. No solamente "Ahora estamos en el fin de semana" sino también "Ahora estoy en la casa de mi abuela", "Ahora hay un peligro evidente de ser atacado por un depredador que se acerca desde el Nornoreste". Esos sistemas de representación tácita no necesitarían *términos* para ser "traducidos por" los distintos términos en las tentativas de los teóricos por captar la información representada tácitamente por esos estados (tales como las tentativas que aparecen encomilladas en la oración anterior). ¡El punto fundamental de la representación tácita es que es tácita! Los estados de un sistema así obtienen sus propiedades semánticas directamente y sólo de sus papeles funcionales globalmente definidos.

A medida que el número de estados diferentes posibles (cada uno con su conjunto distintivo de reglas tácitamente representadas) aumenta, a medida que un sistema se vuelve lo bastante versátil como para ser llevado a una gran cantidad de estados de control significativamente diferentes, este libertinaje exige que el diseño dependa, de una manera u otra, de las economías logradas por la vía del uso múltiple de los recursos. Por ejemplo, allí donde los distintos estados son variaciones sobre un tema (implicando sólo cambios menores en las reglas tácitamente seguidas) se vuelve útil —virtualmente obligatorio— diseñar todo el sistema para cambiar de estado no pasando el control de un subsistema físicamente distinto a otro subsistema casi idéntico, sino cambiando una o más características del subsistema actual, dejando intacto el resto, cambiando los estados por medio de *correcciones y repasos*, digamos, en lugar de *descartar* y *reemplazar*. Las economías de esta clase exigen sistematicidad; los lugares en los que se pueden hacer las sustituciones tienen que tener maneras estables de cambiar sus funciones como una función de la identidad de los sustitutos. De esta manera, todo el sistema empieza a parecerse *de algún modo* a un lenguaje, con contrafiguras para todas las características sintácticas mencionadas al principio.

¿Deberíamos decir que un sistema así emerge finalmente como un sistema de representación interna verdaderamente explícito? Con seguridad habrá en qué usar ese punto de vista de esos sistemas. Pero es importante observar que los elementos "sintácticos" de éstos han de ser considerados primero como poseedores de una semántica totalmente interna "que se refiere" a los discursos de la memoria, las operaciones internas, otros estados del sis-

tema y así sucesivamente y no a cosas y hechos del mundo exterior.[3] Si estos componentes del estado interno se interaniman con habilidad, los estados factibles para ellos pueden tener relaciones delicadamente informativas con los sucesos y cosas del mundo exterior, pero entonces, en lo que respecta a que los estados de esos sistemas se pueden interpretar como poseedores de propiedades semánticas externas, ellos obtienen sus propiedades semánticas exactamente de la misma manera —por exactamente las mismas razones— que los estados representantes meramente tácitos. Es solamente el papel globalmente definido de un estado así (el papel al que se caracteriza en términos de las reglas de funcionamiento que todo el sistema "sigue" cuando pasa a ese estado) lo que fija sus propiedades semánticas informativas o externas.

Por tanto, en un uso extendido, tendría algunas ventajas concederles propiedades semánticas externas indirectas a alguno de los elementos de un sistema tal de estados. Así se podría descubrir que el estado de alguien responsable por su creencia de que la nieve es blanca tiene componentes identificables como el componente "nieve" y el componente "blanco". Pero esas "oraciones del lenguaje del pensamiento", si decidimos que es prudente llamarlas así, están en un constraste llamativo con las oraciones de los lenguajes naturales. Dado lo que las oraciones en castellano "la nieve es blanca" y "la nieve es fría" y "la leche es blanca" significan, podemos decir lo que significa "la leche es fría" (se dé cuenta o no algún orador o integrante del público); pero dados los casos de contrafiguras en el lenguaje mental no podremos *decir* lo que significa el estado "la leche es fría" —puede no significar absolutamente nada— hasta que hayamos determinado su papel global.

Mi objetivo en este ensayo ha sido simplemente explorar —y tal vez mejorar nuestro punto de vista de —algún territorio abierto, empíricamente investigable. Ryle dijo, apriorísticamente, que *no podríamos* ser manipuladores de la representación mental. Fodor y otros han dicho, apriorísticamente, que *debemos* serlo. Algunos de los detalles de la metáfora del ordenador sugieren lo que *podríamos* ser, y al hacerlo así pueden arrojar eventualmente alguna luz sobre lo que somos.

[3] En "*Tom Swift and his Procedural Grandmother*" (1981c) Fodor vé claramente que la semántica interna de los lenguajes programadores y sus afines no resuelven *en sí mismos* el problema de la referencia mental o la intencionalidad. Aquí mi afirmación es que todavía no se nos han dado razones compulsivas para suponer que alguno de los elementos "sintácticos" de los estados internos que tienen verdaderamente propiedades semánticas externas, admitirán alguna interpretación semántica externa directa.

Reflexiones: El lenguaje del pensamiento reconsiderado

La idea de un lenguaje del pensamiento es muy antigua. En su aspecto más reciente, provocativamente patrocinado por Fodor (1975), ha sido objeto de más de una década de escudriñamiento como el eslogan teórico dominante de la ciencia cognitiva. Este es un buen momento para repasar sus perspectivas y ver qué alternativas podrían ser discernibles. Comenzó su carrera reciente como una linda estocada en la oscuridad, inspirada por argumentos relativamente apriorísticos acerca de la única manera en que se podrían satisfacer las exigencias de la cognición más que por medio de ninguna pista empírica muy compulsiva. A medida que se arrojó más luz sobre los mecanismos y métodos de distintos sistemas psicológicos y nerviosos está emergiendo ni reivindicado ni completamente desacreditado. [4]

Habría sido reivindicado con bombos y platillos si hubiera llevado a hipótesis detalladas y verificables acerca de la organización de los procesos cognitivos, pero no ha ocurrido nada parecido. No obstante, nos puede ayudar a entender el estado actual del arte, detenernos a imaginar cómo hubiera sido ese triunfo.

Recordemos los tres niveles de explicación de Marr en la ciencia cognitiva (véase más atrás, págs. 72-73). Podemos describir la reivindicación imaginaria de la hipótesis del lenguaje del pensamiento en términos de una cascada triunfante a través de los tres niveles de Marr.

Supongamos entonces que gracias a los esfuerzos de las fuerzas de tareas de la actiud proposicional, los principios de los cálculos de la psicología popular han llegado a ser especificados tan rigurosamente como los de la aritmética. Y supongamos que estos principios han demostrado funcionar tan bien como la aritmética funciona como un modelo de competencia para las calculadoras. Tendríamos el nivel computacional bien controlado.

Por tanto, a medida que descendemos de este nivel de atribución de creencia puro e infinitamente extensible (en el que, por ejemplo, notamos la diferencia entre creer que un tercio es 0,33333333333 y creer que es 0,333333333333), hasta el infinito, aproximándonos al nivel del algoritmo, supongamos que no tuviéramos que cambiar radicalmente de categorías; los estados reales —extraídos de un conjunto finito de estados posibles— a través del cual pasan los algoritmos tienen un parecido notable con los estados a los que nos referimos en el nivel computacional. Y, finalmente, supongamos que el trazado cartográfico de los algoritmos llevado al hardware nervioso ocurriera sin problemas; descubrimos la ubicación, duración y otros parámetros físicos de las manipulaciones reales de los símbolos postulados en el nivel algorítmico.

Supongamos que todo esto fuera así; entonces Fodor y los demás Realistas estarían en la gloria, puesto que esto sería la reducción de la psicología de

[4] Estas reflexiones resumen los temas desarrollados en recientes trabajos, reseñas y comentarios que consideré demasiado especializados como para ser incluidos en este volumen: 1984a, b, c; 1986c; de próxima aparición e.

la actitud proposicional al Computacionalismo High Church (Dennett, 1984b, 1986e). Habríamos confirmado que las creencias y las otras actitudes proposicionales, son perfectamente reales. Tener una actitud proposicional resulta ser, tal como Fodor originariamente lo expresó, "estar en alguna relación *computacional* con una representación interna" (1975, pág. 198).

Pero supongamos ahora que no resulta así. Supongamos que algo bastante grande tiene que ceder a medida que nos movemos de nuestro modelo idealizado a nivel computacional —que a su vez resulta no ser riguroso— pasando a través del algoritmo hacia el hardware. Esta no es una suposición ociosa, puesto como Fodor mismo lo observa, el gran progreso en la ciencia cognitiva ha sido descubrir los modos de acción de los sistemas motores, perceptivos y sensoriales periféricos, y aquí el progreso ha sido en gran parte una cuestión de demostrar cómo esos sistemas pueden hacer su trabajo *sin* recurrir a ningún nivel de computación que implique representaciones explícitas (Fodor, 1983, véanse también Dennett, 1984a; y Akins, inédito).

Si no nos hace falta, probablemente, un lenguaje del pensamiento, entonces, para ir y venir de las periferias, ¿necesitamos un lenguaje del pensamiento para manejar los procesos de control más centrales de los animales más evolucionados (o simplemente los seres humanos)? Esas "funciones cognitivas más elevadas" como hacer planes, resolver problemas y "fijar creencias" son intuitivamente las funciones que implican *pensar* (como opuesta a "meramente" percibir y actuar). Es aquí, si en alguna parte, que debería triunfar la visión del intelectual. Aquí podríamos muy bien esperar todavía procesos de inferencia en los que "los postulados... están internamente representados y etiológicamente comprometidos". (Fodor, 1981, pág. 120). Pero en realidad casi no ha habido ningún adelanto acerca de la teoría empírica de estas actividades psicológicas más centrales. Fodor paladea la tristeza:

> Para expresarlo en forma contundente, no tenemos ningún formalismo computacional que nos muestre cómo hacer esto, y no tenemos idea de cómo podrían desarrollarse esos formalismos... Si alguien, un Dreyfus, por ejemplo, nos preguntara por qué deberíamos suponer siquiera que la computadora digital es un mecanismo plausible para la simulación de los procesos globales cognitivos, el silencio de la respuesta sería ensordecedor (1983, pág. 129).

El problema no es que los modelos oracionalistas de ese pensamiento no conduzcan a hipótesis plausibles y comprobables. Peor aun, parecen bajar sistemáticamente a callejones sin salida reconocibles; desesperadamente frágiles, ineficientes, y a monstruosidades inflexibles de la ingeniería que apenas podrían guiar a un insecto incólume por la vida. En el centro de las dificultades está el Problema del Armazón de la inteligencia artificial, como lo admite Fodor (1983, págs. 112 y sigs.), que ha probado ser tan resistente a la solución mediante las técnicas ortodoxas de la ciencia cognitiva que se puede afirmar con firmeza que significaría un desastre para las actitudes proposicionales como relaciones computacionales con las representaciones internas (Dennett, 1984c).

La idea de un lenguaje del pensamiento quedaría completamente desacreditada si se hubiera formulado una alternativa clara para ellas que hu-

biera probado que podía manejar las tareas para las que estaba postulada. Fodor arrojó el guante en 1975, reconociendo que la idea de un lenguaje del pensamiento era dura de tragar pero desafiando a los escépticos a encontrar una alternativa. El autor de su epígrafe fue Lyndon B. Johnson: "Soy el único Presidente que tiene" (pág. 27). La ola reciente de modelos conexionistas (McClelland y Rumelhart, 1986) ha sido vista por algunos entusiastas como precisamente esa alternativa, pero a pesar de lo atractivo que es el conexionismo como respuesta al desafío de Fodor, todavía es muy pronto para formular veredictos. Una descripción resumida de esta tendencia reciente (tomada de Dennett, 1984b, 1986e) puede servir, sin embargo, como para dar al menos una impresión de *cómo sería* una alternativa para el lenguaje del pensamiento.

Las cadenas conexionistas desarrolladas hasta ahora son (en el mejor de los casos) *fragmentos* de sistemas cognitivos compuestos por unidades relativamente simples, ricamente interconectadas. Una cadena así difiere comúnmente de los modelos tradicionales (computacionalista de High Church) de la ciencia cognitiva porque tiene

1) memoria y procesamiento "distribuidos", en los que las unidades juegan papeles múltiples, drásticamente equívocos, y en los que la desambiguación se produce sólo "globalmente" (en pocas palabras, no hay "proposiciones" localizadas en las "direcciones" de la memoria);

2) ningún control central sino más bien un sistema parcialmente anárquico de elementos algo competitivos;

3) ningún paso de mensajes complejo entre las unidades;

4) una confianza en las propiedades estadísticas de los conjuntos para lograr efectos;

5) el hacer y deshacer de manera relativamente negligente y poco eficiente los muchos caminos y soluciones hasta que el sistema se consolide después de un tiempo, no necesariamente sobre una respuesta "correcta" predesignada.

Los modelos siguen siendo computacionales en un sentido: están implementados sobre ordenadores y el comportamiento de cada nodo o unidad es una función claramente definida (y computada) del comportamiento de (algunos de) los otros nodos. ¿Cómo difieren entonces estos modelos tan llamativamente en calidad de los modelos tradicionales de la ciencia cognitiva?

Para empezar, el nivel en el cual el modelo es computacional está mucho más cerca de la neurociencia que de la psicología. *Lo que está computarizado* no es (por ejemplo) una implicación de una "proposición" de cálculo del predicado *acerca de Chicago* o una descripción formal *de una transformación gramatical*, sino (por ejemplo) el nuevo valor de algún parámetro parecido a un umbral de algún elemento *que por sí mismo no tiene ningún papel semántico unívoco del mundo externo*. En un nivel tan bajo de descripción, la semántica del medio simbólico de computación se refiere únicamente a los sucesos, procesos, estados, direcciones dentro del cerebro, dentro del sistema computacional mismo.

¿Cómo podríamos hacer entonces que ocurra algo en un sistema así que sea propiamente *acerca de Chicago*? Debe haber indudablemente un nivel

más alto de descripción al que poder atribuirle propiedades semánticas externas ante las características relativamente globales de las actividades de la cadena, pero en un nivel así las interacciones y relaciones entre los elementos semánticos no son computacionales sino —y aquí nos deslizamos temporalmente en la metáfora y la gesticulación— estadísticas, emergentes, holísticas. El mecanismo virtual reconocidamente psicológico en su actividad no es *un mecanismo* en un sentido conocido: su comportamiento no se puede especificar formalmente *en el nivel de vocabulario psicológico* como la computadorización de algún algoritmo de alto nivel. De este modo, en esta visión, el nivel algorítmico es *diferente*, de manera importante, del lenguaje de un mecanismo normal en que no existe ninguna suposición de una traducción directa o relación de implementación entre los fenómenos de alto nivel que sí tienen una semántica del mundo exterior y los fenómenos al nivel algorítmico bajo. Si existiera, el precepto metodológico habitual de la ciencia de la ordenación funcionaría bien. Pasemos por alto el hardware, puesto que las idiosincrasias de su implementación particular no le agregan nada al fenómeno, siempre que el fenómeno sea descrito rigurosamente en el nivel más alto. En los modelos conexionistas, el *hardware* (comúnmente estimulado) sí agrega algo: simplemente cuáles efectos de contenido relativo ocurren realmente (algo que se puede describir sólo estadísticamente en el alto nivel) depende de las características de bajo nivel de la historia del funcionamiento. Las diferentes peculiaridades de la cognición emergen de la actividad, sin estar diseñadas específicamente como para emerger.

Esas cadenas conexionistas han probado ser capaces de realizar una variedad de subtareas cognitivas que hasta ahora se suponía que exigían una maquinaria computacional complicada y basada en reglas, todo sin las representaciones de reglas explícitamente designadas. Para citar quizás el ejemplo más accesible la charla en cadena de Sejnowski (1986) "aprende" a pronunciar el inglés escrito, empezando por un conjunto de disposiciones casual para pronunciar los elementos del texto y ser "corregido" a medida que avanza. Pronto exhibe exactamente el tipo de conducta que ha sido considerado hasta ahora como sintomático de la computación basada en reglas: generaliza a partir de lo que ha visto, pero llega a reconocer excepciones a sus generalizaciones, y generaliza sobre esas excepciones. Sin embargo, en ninguna parte de la charla en CADENA hay reglas *explícitamente* representadas; están, por supuesto, representadas *tácitamente* en la estructura dispositiva emergente de la cadena, pero en ésta no ocurre nada que se parezca *a verificar para ver si una regla es pertinente* o *a buscar el término fijado en una tabla de excepciones*, paradigmas de las actividades intelectualistas ocultas que Ryle desacreditó.

Todavía hay muchos fundamentos para el escepticismo acerca del conexionismo. Todas las cadenas existentes logran sus notables efectos con la ayuda de puntales ominosamente irreales, tales como el "maestro" que ayuda a meter correcciones en la charla en CADENA, y, a un nivel más bajo, variedades de un tráfico en ambas direcciones entre los nodos que hasta ahora son desconocidos para la neurociencia. Tal como las primeras falsas esperanzas en inteligencia artificial (véase *Brainstorms*, cap. 7) existen preocupa-

ciones acerca de *"escalar"*: los modelos que trabajan en forma impresionante, al ser restringidos a unas pocas docenas o cientos de elementos a veces se detienen en seco en sus carriles cuando se los expande a dimensiones realistas. Los caminos plausibles para escapar de estos problemas ya han sido identificados, pero llevará varios años de exploración ver dónde llevan.

Si la ciencia cognitiva fuera la ciencia aeronáutica, describiríamos la situación actual de la siguiente manera. El primer gran paso en la tentativa de crear Máquinas Voladoras fue la insistencia por parte de los fundadores de que los llamados místicos al *tejido maravilla* (Dennett, 1984c) no tenían ninguna esperanza: los diseños tenían que basarse en principios mecánicos sólidos y bien comprendidos. Siguieron años de esfuerzos en pos de brillantes diseños, y una gran variedad de máquinas promisorias —pero desafortunadas— se construyeron con los materiales bien conocidos entonces: ladrillos, cemento y madera. Nada voló realmente. Luego llegaron los conexionistas con lo que parecía ser una variedad de nuevos tejidos maravilla sintéticos. Todavía está por demostrarse cómo estas telas excitantes se pueden coser hasta transformarse en Máquinas Voladoras enteras que funcionan. Tal vez parte del primer trabajo de diseño se pueda adaptar a los nuevos materiales. [5]

McClelland y Rumelhart (1986) y Smolensky (de próxima aparición) admitieron los problemas todavía no encarados por los conexionistas. Uno de ellos sigue siendo el problema de la generatividad. El mejor argumento en favor del lenguaje del pensamiento es todavía la afirmación de que *de algún modo* todo el sistema, al menos en los seres humanos, debe ser capaz de una capacidad virtualmente ilimitada para diferenciar y luego basar su conducta en rasgos de esas diferenciaciones. Parafraseando el grito de batalla de Fodor, no hay diferenciación sin representación. Todavía tenemos un solo ejemplo claro de un método de representación *infinitamente extensible y articulado*: un lenguaje natural. El sistema de generatividad de cualquier cosa que sirva para representar al mundo en nosotros será así casi seguramente interpretable retrospectivamente como si estuviera compuesto de términos y oraciones en un lenguaje, pero eso no reivindicaría en sí mismo la idea de los objetos lingüísticos como las estructuras básicas de la cognición. Puede muy bien ser que este nivel lingüístico aparezca como una propiedad inocentemente emergente de las actividades distribuidas de otras unidades más que la manera en que es discernible un cuadro en un mosaico compuesto por elementos que se sostienen por principios que desconocen los límites de los elementos del cuadro.

¿Cuán estrechamente unidos están los destinos de la hipótesis del lenguaje del pensamiento por un lado y el realismo con respecto a las actitudes proposicionales por el otro? Fodor siempre sostuvo que se mantienen en pie o caen juntos y muchos otros lo han acompañado, incluyendo Stich y los Churchland. En la eventualidad de que nuestro modelo de competencia de psicología popular sea para siempre sólo un modelo de competencia idealizado que guarda poca relación con los mecanismos subyacentes —y esto sería

[5] Para un esbozo pionero, véase, por ejemplo, la demostración de Touretzky y Hinton (1985) acerca de cómo armar un *sistema de producción* (una de las arquitecturas del computacionalismo High Church) a partir de una máquina Boltzmann (una de las nuevas estructuras conexionistas).

así si el conexionismo triunfara— ¿qué deberíamos decir al final acerca de las categorías de la psicología popular? Stich (1983, págs. 221-28) que todavía encuentra atractiva la idea de un lenguaje del pensamiento, propone alguna esperanza en favor de una "perspectiva panglossiana modificada" en la que la psicología madura "se adhiere muy estrechamente al patrón presupuesto por la psicología popular", pero "en momentos más pesimistas" anticipa justamente este resultado negativo. Del mismo modo pero más animosamente lo hace Churchland (1981). Ambos declaran que esto demostraría que después de todo no hay estados tales como las creencias; anunciarían la muerte de la psicología popular y todas sus versiones emperifolladas como la psicología de la actitud proposicional.

¿Qué pondrían en su lugar? Cada vez que a Stich le abruma el optimismo y anticipa que Fodor demostrará estar *casi* en lo cierto, reemplazaría la psicología popular por una teoría puramente "sintáctica"; se aferraría a las versiones de las categorías populares —los estados de creencia, por ejemplo— pero las desinterpretaría. En sus momentos pesimistas se uniría a Churchland y haría exactamente lo contrario: se aferraría a la plenitud del contenido o intencionalidad de una u otra cosa interna pero abandonaría la presunción de que estos vehículos internos del significado se comportaban sintácticamente, en las formas en que los realistas han supuesto que se comportan las creencias. Las creencias privadas de contenido versus el contenido privado de creencias.

Se puede decir algo en favor de cada una de estas posiciones. ¡Por cierto que se las podría "reivindicar" inmediatamente a las dos por medio de triunfos futuros en la psicología empírica! Puesto que si tengo razón, hay en realidad dos clases de fenómenos a los que la teoría de la filosofía popular acerca de las creencias alude de manera confusa: los estados verbalmente contaminados pero sólo problemática y derivativamente plenos de contenido de los usuarios del lenguaje ("opiniones") y los estados más profundos de lo que se podría llamar creencia animal (las ramas y los perros tienen creencias pero ninguna opinión). Stich, con su teoría sintáctica, puede estar recomendando más o menos la manera correcta de manejarse con las opiniones, mientras que Churchland puede ser el gurú metodológico para los que teorizan acerca de los otros, los guías de conducta interna, los estados sensibles a la información, que no son para nada "oracionales" en su estructura.

Sin embargo elijo no seguir ninguna de las dos ramas eliminatorias. Mi incapacidad para unirme a alguno de los dos campos no se trata, como la inmovilidad del asno de Buridan, de que considere de la misma manera los atractivos de ambos y sea incapaz de elegir. Puesto que aunque reconozco estos atractivos veo un problema compartido en su eliminatoriedad extrema: ¿qué le dirán al juez hasta que el resto del mundo los alcance y comparta su visión del mundo? Es decir, cuando se los llame a testimoniar bajo juramento en un juzgado de justicia y el juez les pregunte si *creen* que han visto al acusado alguna vez, ¿qué dirán? Seguramente deben negar que están diciendo lo que creen, puesto que creen (a-já) que no existe tal cosa como la creencia. Es decir que son *de la opinión* (¿servirá eso?) de que la creencia no existe. Lo que quieren decir es, la teoría que ellos defienden o apoyan, este... no tiene lugar para las creencias en su ontología.

Esta es una línea de ataque conocida y no es ninguna novedad para Stich y Churchland. La clase anterior de eliminatoriedad de Skinner ha sido a menudo proclamada como sistemáticamente autoimpugnante por dichas razones. Estoy seguro de que ninguno de los argumentos *a priori* que aparentan presentar estas categorías de la imagen manifiesta como inmunes al descrédito científico es sólido, y no apruebo ninguno de ellos. Entiendo que el problema no es directamente doctrinario sino táctico (véase "Quining Qualia" de próxima aparición d y el capítulo 1 de *Content and Consciousness*). Cuando digo que la elección es táctica no estoy llamando simplemente a la discreción la mayor parte del valor y aconsejando una retirada decorosa ante los grandes batallones del sentido común enfurecido. Eso sería cobarde y deshonesto (aunque se podría defender muy bien el juzgado: ¿qué prefiere usted, que se le haga justicia al acusado o perder un día precioso en el juzgado fastidiando al juez con una lección de filosofía inconvincente?). Tácticamente, es de mayor importancia para el filósofo de la ciencia no perder de vista el tremendo —si bien defectuoso— poder predictivo de la actitud intencional. El juez puede ser una figura ceremonial, pero no es un médico brujo; su deseo oficial de saber lo que usted cree no es irracional. Su método es, más allá de toda duda, la mejor manera de llegar a la verdad que conocemos. ¿Hay alguna razón para creer que en la Edad de Oro de la Psicosintaxis Eliminatoria o la Neurobiología Eliminatoria surgirá algún método rival serio de búsqueda de la verdad?

Churchland ofrece especulaciones optimistas acerca de este mismo punto, y fuera de todo lo que alguien pudiera decir en este momento, podría resultar profético. Pero mientras tanto tiene que seguir algún lineamiento acerca del status de los poderes (aparentemente) predictivos y explicativos no sólo de la humilde psicología popular, sino también de las ciencias sociales académicas. ¿Qué está sucediendo en los modelos académicos que presuponen agentes racionales? ¿Cómo puede explicar el poder de los psicólogos cognitivos para diseñar experimentos fructíferos que exigen suposiciones acerca de las creencias que sus sujetos tienen acerca de la situación de ensayo, del deseo que han inculcado en sus sujetos de prestar atención al mensaje en el oído izquierdo, etcétera? (Dennett, 1985c). Ese poder es perfectamente real y exige una explicación. ¿Podría decir que, en espera de la Edad de Oro, se dispone de una especie de cálculo instrumentalista, de origen humilde pero también de cierta efectividad, no es cierto?

Supongamos, por el bien del drama, que resulte que la psicología cognitiva subpersonal de algunas personas sea dramáticamente diferente de la de otras. Uno puede imaginar los titulares de los diarios: "Los científicos prueban que la mayoría de los zurdos son incapaces de tener creencias" o "Asombroso descubrimiento: los diabéticos no tienen ningún deseo". Pero esto no es lo que diríamos, como sea que resultara la ciencia.

Y nuestra renuencia no sería sólo conservadurismo conceptual, sino el reconocimiento de un hecho empírico evidente. Puesto que aunque debamos dejar que zurdos y diestros (u hombres y mujeres o cualquier otro subgrupo de personas) sean todo lo internamente diferentes que se quiera, ya sabemos que hay modelos fuertes, confiables de los que participan todas las personas

de conducta normal: los modelos que describimos tradicionalmente en términos de creencia y deseo y los otros términos de la psicología. ¿Qué se expandió por el mundo el 20 de julio de 1969? La creencia de que un hombre había pisado la Luna. El efecto de recibir esa información no fue la misma ni siquiera en dos personas, y los caminos causales que ocurrían en el estado que todos tenían en común fueron sin duda igualmente variados, pero la afirmación de que por tanto nadie tenía nada en común —nada de importancia en común— es evidentemente falso. Hay una cantidad infinita de maneras en las que se podrían distinguir en forma confiable las que tienen la creencia de las que no la tienen, y habría una elevada correlación entre los métodos. La ciencia no debería —ni podría— darle la espalda a algo así.

¿Cómo veo entonces la Edad de Oro? En forma muy parecida a la de Churchland, con algunos cambios en el énfasis. En el capítulo 3, desplegué los tipos. Primero estará nuestra vieja y confiable amiga, la psicología popular y, segundo, su idealización conscientemente abstracta: la teoría del sistema intencional. Finalmente habrá una teoría bien confirmada a un nivel entre la psicología popular y la biología desnuda, la psicología cognitiva subpersonal. Ahora podemos decir algo más acerca de cómo sería: será "cognitiva" en el sentido de que describirá procesos de transformación de información entre los items cargados de contenido —representaciones mentales— pero su estilo no será "computacional", los items no parecerán ni se comportarán como oraciones manipuladas en un lenguaje del pensamiento.

7

Los sistemas intencionales en la etología cognitiva: Defensa del "Paradigma panglossiano"*

El problema

El campo de la etología cognitiva proporciona una rica fuente de material para el análisis filosófico del significado y la mentalidad y hasta ofrece algunas perspectivas tentadoras para que los filósofos contribuyan bastante directamente al desarrollo de los conceptos y métodos de otra disciplina. Como filósofo, un advenedizo con sólo un conocimiento superficial del campo de la etología, noto que los nuevos etólogos, después de haber desechado el chaleco de fuerza del conductismo y de haberse quitado con fastidio sus pesados chanclos, están buscando a su alrededor con cierta inseguridad algo presentable para ponerse. Están buscando un vocabulario teórico que sea poderosamente descriptivo de los datos que están destapando y al mismo tiempo un método teóricamente fructífero de formular hipótesis que lleven, *eventualmente*, a los modelos de procesamiento de información del sistema nervioso de los seres que están estudiando (véase Roitblat, 1982). Hay un largo camino desde la observación de la conducta, digamos, de los primates en estado salvaje hasta la ratificación de los modelos neurofisiológicos de su actividad cerebral, y encontrar una manera provisional sólida de expresarlo no es una tarea trivial. Puesto que me parece que los problemas metodológicos y conceptuales que los etólogos deben enfrentar tienen un parecido notable con los problemas que yo y otros filósofos hemos estado tratando de resolver recientemente, me siento tentado a entremeterme y ofrecer, primero, un rápido análisis del problema; segundo, una propuesta para tratarlo (que llamo la teoría del sistema intencional); tercero, un análisis de la continuidad de la teoría del sistema intencional con la estrategia o actitud teórica en una teoría evolutiva a menudo llamada *adaptacionismo*; y finalmente, una defensa limitada del adaptacionismo (y su prima, la teoría del sistema intencional) en contra de las críticas recientes de parte de Stephen J. Gould y Richard C. Lewontin.

* Publicado originalmente en *The Behavioral and Brain Sciences* 6 (1983): 343-90 y vuelto a imprimir con autorización.

La metodología de la filosofía, tal como es, incluye como una de sus estrategias más populares (y a menudo auténticamente fructíferas) la descripción y el estudio de situaciones completamente imaginarias, complicados experimentos del pensamiento que aíslan para su escrutinio las características presumiblemente críticas en algún terreno conceptual. En *Word and Object*. W.V.O. Quine (1960) nos brindó un amplio estudio de las tareas probatorias y teóricas que enfrentan el "traductor radical", el antropólogo-lingüista imaginario que penetra en una comunidad completamente ajena —sin intérpretes ni guías bilingües— y que debe deducir usando cualquier método científico de que disponga, el lenguaje de los nativos. De este experimento del pensamiento surgió la tesis de Quine acerca de la "indeterminación de la traducción radical", la afirmación de que siempre debe ser posible en principio producir manuales diferentes de traducción importantes, igualmente bien apoyados por todas las pruebas, para cualquier lenguaje. Uno de los rasgos más polémicos de la posición de Quine a través de los años han sido sus escrúpulos intransigentemente conductistas acerca de cómo caracterizar la tarea que enfrenta el traductor radical. ¿Qué pasa con la tarea de la traducción radical cuando se abandona el compromiso con un enfoque y una terminación única de un lenguaje (o una interpretación única de los "estados mentales" de un ser) si uno se permite el vocabulario y los métodos del "cognitivismo"? El tema podría explorarse por la vía de otros experimentos del pensamiento y lo ha sido en algunos aspectos (Jennett, 1976; Dennett, 1971; Lewis, 1974) pero las investigaciones del mundo real de Seyfarth, Cheney y Marler (1980) con los micos de Africa del Sur nos será más útil en esta ocasión. Esos micos forman sociedades de cierta especie, y tienen un lenguaje de cierto tipo y por supuesto que no hay intérpretes bilingües que les presten ninguna ayuda a los traductores radicales del lenguaje mical. Esto es lo que encuentran:

> Los micos producen distintos gritos de alarma ante los distintos depredadores. Las grabaciones de las alarmas que se hicieron escuchar de nuevo cuando los depredadores estaban ausentes, hicieron que los monos corrieran a los árboles a causa de alarmas de leopardos, que buscaran alarmas de águilas y miraran para abajo a causa de alarmas de serpientes. Los monos adultos gritan fundamentalmente ante los leopardos, las águilas imperiales y las víboras pitón pero las crías perciben alarmas de leopardo ante distintos mamíferos, alarmas de águila ante muchos pájaros y alarmas de víbora ante distintos objetos con aspecto de serpiente. La clasificación de los depredadores mejora con la edad y la experiencia. (Extracto de Seyfarth, Cheney y Marler, 1980, pág. 801.)

Este extracto se esconde, como se puede notar, en un lenguaje de la conducta casi puro: el lenguaje de la *ciencia* aun si ya no es más exclusivamente eso. Es lo bastante informativo como para ser provocador. ¿Cuánto lenguaje, uno quisiera saber, tienen en realidad los micos? ¿Se comunican *verdaderamente*? *¿Quieren decir lo que dicen?* ¿Exactamente qué interpretación les podemos dar a estas actividades? ¿Qué nos dicen en realidad estos datos acerca de las capacidades cognitivas de los micos? ¿De qué manera son

—deben ser— ellas como capacidades cognitivas humanas, y en qué modos y hasta qué grado son los micos más inteligentes que otras especies en virtud de estos talentos "lingüisticos"? Estas preguntas recargadas —las más naturales de formular en estas circunstancias— no caen exactamente en el campo de ninguna ciencia, pero sean o no las preguntas correctas que los científicos deben formular, son con seguridad las que todos nosotros, como seres humanos fascinados que nos enteramos de la aparente semejanza de los micos con nosotros, queremos que sean contestadas.

El cognitivista querría sucumbir a la tentación de usar el lenguaje mentalista común más o menos a su valor nominal y responder directamente preguntas tales como: ¿Qué *saben* los monos? ¿Qué *quieren*, *entienden* e *intentan decir*? Al mismo tiempo, el punto primordial de la investigación de los cognitivistas no es satisfacer la curiosidad del lego acerca del CI, por así decirlo, de sus primos los simios, sino trazar un gráfico de los *talentos* de estos animales de paso que trazan un cuadro de los *procesos* cognitivos que explican esos talentos. ¿Podría el lenguaje cotidiano de la creencia, el deseo, la esperanza, el reconocimiento, la comprensión y otras cosas por el estilo, ser útil también como el lenguaje abstracto adecuadamente riguroso en el que describir las competencias cognitivas?

Yo sostendré que la respuesta es sí. Sí, si tenemos cuidado con lo que estamos haciendo y diciendo cuando usamos palabras comunes "creer" y "querer", y comprendemos las presunciones e implicaciones de la estrategia que debemos adoptar cuando usamos estas palabras.

La decisión de dirigir la ciencia propia en términos de creencias, deseos y otras nociones "mentalistas", la decisión de adoptar la "actitud intencional", no es una clase de decisión desusada en la ciencia. La estrategia básica de la que éste es un caso especial es conocida: cambiar los niveles de explicación y descripción para ganar acceso a un poder predictivo o generalidad mayores —adquirida, comúnmente, al coste de sumergir el detalle y galantear la trivialización por un lado y la falsificación fácil por el otro. Cuando los biólogos que están estudiando alguna especie eligen llamar *alimento* a algo en el entorno de esa especie y lo dejan así, pasan por alto los detalles tramposos de la química y la física de la nutrición, la biología de la masticación, digestión, excreción y el resto. Aun suponiendo que los detalles de esta biología más refinada sean todavía mal comprendidos, la decisión de saltar hacia adelante, anticipándose a la biología de grano fino, y confiar en el buen comportamiento del concepto de alimento al nivel de la teoría apropiada a él recibirá probablemente la aprobación de los corredores de riesgos más conservadores.

La decisión de adoptar la actitud intencional es más arriesgada. Cuenta con la solidez de algún concepto de información todavía no descrito de manera precisa, no el concepto legitimado por la teoría de información Shannon-Weaver (Shannon, 1949) sino más bien el concepto de lo que a menudo se llama *información semántica*. Una manera más o menos estándar de presentar la distinción todavía imperfectamente entendida entre estos dos conceptos de información, es decir que la teoría Shannon-Weaver mide la *capacidad* de transmisión de la información y la de los vehículos de almace-

naje de ésta, pero se mantiene muda acerca de los *contenidos* de esos canales y vehículos, que serán el tema de la teoría aun por ser formulada de la información semántica (véase Dretske, 1981, para una tentativa por llenar el vacío entre los dos conceptos). La información desde el punto de vista semántico es un producto perfectamente real pero muy abstracto, cuyo almacenaje, transmisión y transformación se narra informalmente —pero de manera certera— en la conversación común en términos de creencias y deseos y los otros estados y actos que los filósofos llaman *intencionales*.

La teoría del sistema intencional

La intencionalidad, en la jerga filosófica, es —en una palabra— *acerquidad*. Algunas de las cosas, estados y sucesos del mundo tienen la interesante propiedad de *ser acerca* de otras cosas, estados y sucesos; figurativamente, señalan otras cosas. Esta flecha de referencia o acerquidad ha estado sometida a un intenso escrutinio filosófico y ha engendrado mucha polémica. Para nuestros propósitos podemos extraer delicadamente dos puntos de este caldero hirviente, simplificarlos exageradamente y pasar por alto aspectos importantes que son tangenciales con nuestros intereses.

Primero, podemos observar la presencia de la intencionalidad —acerquidad— como tema de nuestras discusiones, señalando la presencia de un rasgo *lógico* peculiar de toda discusión así. Las oraciones que atribuyen estados o sucesos intencionales a los sistemas utilizan modismos que exhiben *opacidad referencial*: introducen cláusulas en las que la regla de sustitución normal, permisiva, no se sostiene. Esta regla es simplemente la codificación lógica de una máxima que dice que una rosa olería con la misma dulzura si tuviera cualquier otro nombre. La regla dice que si se tiene una oración verdadera y se la altera reemplazando un término de ella por otro, un término diferente que todavía se refiere a la misma cosa o cosas, la oración nueva también será verdadera. Lo mismo ocurre con las oraciones falsas, meramente cambiando los medios de elegir los objetos acerca de los cuales la oración no puede convertir una mentira en una verdad. Por ejemplo, supongamos que "Bill es el chico mayor de la clase"; entonces, si es verdad que

1) Mary está sentada junto a Bill,

por tanto, si sustituimos "el chico mayor de la clase" por "Bill" obtenemos

2) Mary está sentada al lado del chico mayor de la clase,

que *debe* ser verdad si la otra oración lo es.

Una oración que tiene un *modismo intencional* contiene, sin embargo, una cláusula en la cual esa sustitución puede convertir la verdad en mentira y viceversa. (A este fenómeno se lo llama *opacidad referencial* porque los términos de esa cláusula están escudados o aislados del análisis lógico por una barrera. El análisis lógico normalmente "ve a través" de los términos el mundo al que éstos se refieren.) Por ejemplo, sir Walter Scott escribió *Waverly*, y Bertrand Russell (1905) nos asegura que

3) Jorge IV se preguntaba si Scott era el autor de *Waverly*,

pero ciertamente parece improbable que

4) Jorge IV se preguntara si Scott era Scott.

(Como señala Russell, "es prácticamente imposible atribuirle un interés en la ley de identidad al primer caballero de Europa" [1905, pág. 485].) Para dar otro ejemplo, supongamos que decidimos que es verdad que

5) Burgess teme que el ser que se agita en el arbusto sea una pitón

y supongamos que en realidad el ser que está en el arbusto es Robert Seyfarth. No desearemos extraer la conclusión de que

6) Burgess teme que Robert Seyfarth sea una pitón.

Pues bien, en cierto sentido lo hacemos y en cierto sentido también queremos insistir en que, bastante extrañamente, el rey Jorge se preguntaba si Scott era Scott pero no se lo expresaba así a él mismo, y no es así como Burgess imaginó el ser en el arbusto tampoco, es decir, como Seyfarth. Es el sentido de imaginar *cómo*, ver *cómo*, pensar en *cómo* lo que los modismos intencionales enfocan.

Un ejemplo más: supongamos que usted crea que su vecino de al lado podría ser un buen marido para alguien y suponga que sin usted saberlo él es el Estrangulador Loco. Aunque en un sentido muy forzado se podría decir que usted cree que el Estrangulador Loco sería un buen marido para alguien, en otro sentido más natural usted no lo cree, puesto que hay otra creencia —muy extraña e improbable— que usted seguramente no tiene y que podría llamarse mejor la creencia de que el Estrangulador Loco sería un buen marido.

En esta resistencia a la sustitución, la insistencia de que para *algunos* propósitos, cómo se llama rosa a una rosa tiene mucha importancia, lo que hace que los modismos intencionales sean idealmente adecuados para hablar acerca de las maneras en que la información está representada en la cabeza de las personas y otros animales. De manera que el primer punto acerca de la intencionalidad es simplemente que podemos confiar en un conjunto señalado de modismos para tener esta característica especial de ser sensibles a los *medios de referencia* usados en las cláusulas que ellos introducen. Los más familiares de esos modismos son "cree que", "sabe que", "espera (que)", "quiere (que sea cierto que)", "reconoce (que)", "entiende (que)". En pocas palabras, el vocabulario "mentalista" evitado por los conductistas y celebrado por los cognitivos está muy bien escogido por el test lógico de la opacidad referencial.

El segundo punto a extraer del caldero es algo polémico, aunque tiene muchos partidarios que han llegado aproximadamente a la misma conclusión por distintos caminos: el uso de los modismos intencionales tiene una presunción o suposición de *racionalidad* en el ser o el sistema al que se le atribuyen los estados intencionales. Lo que esto significa se aclarará si nos volvemos ahora a la actitud intencional en relación con los micos de Sudáfrica.

Los micos como sistemas intencionales

Adoptar la actitud intencional hacia estos monos es decidir —tentativamente, por supuesto— tratar de caracterizar, predecir y explicar su conduc-

ta usando modismos intencionales, tales como "cree" y "quiere", una práctica que da por sentado o presupone la racionalidad de los micos. Diremos que un mico es un sistema intencional, algo cuya conducta se puede predecir atribuyéndole creencias y deseos (y por supuesto racionalidad). ¿*Cuáles* creencias y deseos? Hay muchas hipótesis disponibles y son verificables en virtud del requisito de la recionalidad. Primero, observemos que hay distintos grados de sistemas intencionales.

Un sistema intencional *de primer orden* tiene creencias y deseos (etc.) pero ninguna creencia ni deseo *acerca de* las creencias y deseos. De esta manera, todas las atribuciones que le hacemos a un sistema intencional meramente de primer orden tiene la forma lógica de

7) *x* cree que *p*

8) *y* quiere que *q*

donde *p* y *q* son cláusulas que no contienen en sí mismas modismos intencionales. Un sistema intencional de *segundo orden* es más sofisticado: tiene creencias y deseos (y sin duda otros estados intencionales) acerca de las creencias y los deseos (y otros estados intencionales) tanto los de otros como los propios. Por ejemplo,

9) *x quiere* que *y crea* que *x* tiene hambre

10) *x cree que y espera que x* salte hacia la izquierda

11) *x teme* que *y descubra que x* tiene un escondite para la comida

Un sistema intencional de *tercer orden* es el que es capaz de estados como

12) *x quiere* que *y crea* que *x cree* que está solo

Un sistema de cuarto orden *querría* que usted *pensara* que él *entendía* que usted le estaba *pidiendo* que se fuera. ¿Cuán alto podemos llegar los seres humanos? En principio, sin duda, hasta el infinito, pero en realidad sospecho que usted se pregunta si yo me doy cuenta de lo difícil que es para usted estar seguro de entender si yo quiero decir que usted admite que yo creo que usted quiere que yo explique que la mayor parte de nosotros podemos seguir con atención sólo cinco o seis órdenes, en la mejor de las circunstancias. Véase Cargile (1970) para una explotación elegante pero sobria de este fenómeno.

¿Cuán buenos son los micos? ¿Son verdaderamente capaces de una intencionalidad de tercer orden o más? La pregunta es interesante en varios frentes. Primero, estos órdenes ascienden por lo que es *intuitivamente* una escala de inteligencia; las atribuciones de un orden más alto nos parecen mucho más sofisticadas, mucho más humanas y que exigen mucha más inteligencia. Hay varios diagnósticos plausibles de esta intuición. Grice (1957-1969) y otros filósofos (véase especialmente Bennett, 1976) han desarrollado un problema complicado y esmeradamente discutido para el punto de vista de que la *comunicación* auténtica, los hechos del habla en el sentido fuerte, humano de la palabra, dependen de *por lo menos* tres órdenes de la intencionalidad tanto en el hablante como en el auditorio.

No todas las interacciones entre los organismos son comunicativas. Cuando le pego a una mosca no me estoy comunicando con ella ni tampoco si le abro la ventana para que se escape. ¿Se comunica un perro ovejero con las ovejas que cuida? ¿Se comunica un castor al dar coletazos? ¿Y se comunican las abejas al ejecutar sus famosas danzas? ¿Se comunican los bebés humanos

con sus padres? ¿En qué momento se puede estar seguro de estarse comunicando verdaderamente con un bebé? La presencia de signos lingüísticos específicos no parece ni suficiente ni necesaria. (Puedo dar órdenes en inglés para conseguir que mi perro haga cosas, pero ésa es en el mejor de los casos una forma de comunicación descolorida si se la compara con la ligera elevación de la ceja por medio de la cual puedo hacer que alguien sepa que debería cambiar el tema de nuestra conversación.) La teoría de Grice proporciona un marco mejor para contestar estas preguntas. Define criterios para la comunicación *intuitivamente plausibles y formalmente poderosos que implican*, como mínimo, la atribución correcta a los comunicadores de estados intencionales de tercer orden, tales como:

13) El locutor *trata* de que el auditorio *reconozca* que el locutor *trata* de que el auditorio produzca una respuesta *r*.

De modo que una razón para estar interesado en la interpretación intencional de los micos es que promete contestar —o al menos ayudar a contestar— las preguntas: ¿es esta conducta realmente lingüística? ¿Se están comunicando realmente? Otra razón es que el orden más alto es una señal conspicua de las atribuciones sobre las que se especuló en la literatura sociobiológica acerca de esos rasgos interactivos como el altruismo recíproco. Hasta se ha especulado (Trivers, 1971), con que la creciente complejidad de la representación mental exigida para el mantenimiento de sistemas de altruismo recíproco (y otras complejas relaciones sociales) llevó a través de la evolución, a una especie de carrera de armas del poder cerebral. Humphrey (1976) llega a conclusiones similares por un camino diferente y en ciertos aspectos menos especulativo. Puede haber entonces muchos caminos hacia la conclusión de que el orden más alto de la caracterización intencional es una señal profunda —y no simplemente un síntoma confiable— de inteligencia.

(No quiero sugerir que estas órdenes proporcionen una escala uniforme de ningún tipo. Como varios críticos me lo han señalado, la primera reiteración —a un sistema internacional de *segundo orden*— es el paso crucial de la repetición. Una vez que uno tiene en su repertorio el principio de la *incrustación*, la complejidad de lo que puede entonces esperar en cierto sentido parece plausiblemente más una limitación de un espacio de la memoria o de la atención o del "espacio de trabajo cognitivo" que una medida fundamental de la elaboración del sistema y gracias al "fraccionamiento" y otros auxiliares artificiales de la memoria, no parece haber ninguna diferencia *interesante* entre, digamos, un sistema intencional de cuarto orden y uno de quinto. (Véase Cargile, 1970, para mayores reflexiones acerca de los límites naturales de la reiteración.)

Pues bien, volvamos ahora a la pregunta empírica de cuán buenos son los micos. Por el bien de la simplicidad podemos restringir nuestra atención a un solo acto aparentemente comunicativo de un solo mico, Tom, que, supongamos, lanza una llamada de alarma de parte de los leopardos en presencia de otro mico, Sam. Podemos componer ahora un conjunto de interpretaciones intencionales competitivas de esta conducta ordenadas de arriba abajo, de la romántica a la aguafiestas. He aquí una hipótesis (relativamente) romántica (con algunas variaciones a verificar en la cláusula final):

Cuarto orden

Tom *quiere* que Sam *reconozca* que Tom *quiere* que Sam *crea* que
hay un leopardo
hay un cuadrúpedo
hay un animal vivo más grande que una caja de pan.

Una hipótesis menos excitante a confirmar sería esta versión de tercer orden (podría haber otras):

Tercer orden

Tom *quiere* que Sam *crea* que Tom *quiere* que Sam corra hacia los árboles.

Observe que este caso particular de tercer orden difiere del caso de cuarto orden al cambiar la categoría del hecho del habla: según esta lectura la advertencia de leopardos es imperativo (un pedido o una orden), no una declaración (que informa a Sam acerca del leopardo). La importante diferencia entre interpretaciones imperativas y declarativas de las elocuciones (véase Bennett, 1976, secciones 41,51) se puede captar —y entonces pueden explorarse las diferencias de conducta delatoras— en cualquier nivel de descripción por encima del segundo orden, en el cual, ***ex hypothesi***, no hay ninguna intención de pronunciar un acto del habla de ninguna de las dos variedades. Hasta en el segundo orden, sin embargo, se expresa una distinción relativa en el efecto deseado en el auditorio que en principio se puede detectar del punto de vista de la conducta en las siguientes variaciones:

Segundo orden

Tom *quiere* que Sam *crea* que
hay un leopardo
debería correr hacia los árboles.

Esta difiere de las dos anteriores en no suponer que la acción de Tom no implica ("en la mente de Tom") ningún reconocimiento por parte de Sam del papel propio de Tom en la situación. Si Tom pudiera lograr su objetivo igualmente bien gruñendo como un leopardo, o atrayendo de algún modo la atención de Sam hacia el leopardo sin que Sam reconociera la intervención de Tom, éste sería un caso sólo de segundo orden. (Véase *quiero* que usted *crea* que no estoy en mi oficina, de manera que me quedo muy quieto y no contesto su llamada a la puerta. Eso no es comunicarse.)

Primer orden

Tom *quiere* hacer que Sam corra hacia los árboles (y tiene este truco de hacer un ruido que produzca ese efecto. Utiliza el truco para inducir cierta respuesta en Sam).

En esta lectura el grito del leopardo pertenece a la misma categoría de aparecerse detrás de alguien y decir "¡Bu!", no sólo el efecto intentado no de-

pende del reconocimiento de parte de la víctima de la intención del perpetrador; no es necesario que el perpetrador tenga noción alguna de la mente de la víctima; un ruido fuerte detrás de ciertas cosas hace que salten.

Cero orden

Tom (como otros micos) es propenso a tres clases de ansiedad o excitación: ansiedad por los leopardos, por las águilas y por las serpientes.[1] Cada una tiene su vocalización sintomática característica. Los efectos de estas vocalizaciones en otros tienen una tendencia feliz, pero sólo se trata de tropismo, tanto en el que emite como en el auditorio.

Hemos llegado al fondo aguafiestas del tonel; una explicación que no le atribuye al mico ninguna mentalidad, ninguna inteligencia, ninguna comunicación ni ninguna intencionalidad. Hay otras explicaciones posibles. Yo elegí estas candidatas por su simplicidad y vivacidad. El canon de parsimonia de Lloyd Morgan nos ordena decidirnos por la hipotesis más aguafiestas, menos romántica que explica sistemáticamente la conducta observada y observable, y durante mucho tiempo el credo conductista de que se podía hacer que las curvas se adaptaran bien a los datos en el nivel más bajo impidió la exploración de la explicación que se puede dar de las sistematizaciones de un orden más alto, a un nivel más alto de la conducta de esos animales. La afirmación de que *en principio*, siempre se puede contar una historia del orden más bajo de cualquier comportamiento animal (una historia enteramente fisiológica y hasta una historia sobriamente conductista de inimaginable complejidad) ya no interesa más. Es como afirmar que en principio, el concepto de alimento puede ser ignorado por los biólogos —o el concepto de célula o de gen, lo mismo da— o como afirmar que en principio se puede contar una historia puramente en el nivel electrónico sobre el comportamiento de cualquier ordenador. Actualmente estamos interesados en preguntar qué beneficios en perspicuidad, en poder predictivo, en generalización, podrían resultar si adoptáramos una hipótesis de más alto nivel que diera un paso arriesgado en la caracterización intencional.

La cuestión es empírica. La táctica de adoptar la actitud intencional no se trata de reemplazar investigaciones empíricas por investigaciones apriorísticas ("de sillón") sino de usar la actitud para sugerir qué preguntas empíricas brutales hacerle a la naturaleza. Podemos verificar las hipótesis en competencia explotando la presunción de racionalidad de la actitud intencional. Podemos empezar en cualquier extremo del espectro, ya sea tratando de encontrar los tipos deprimentes de evidencia que *rebajarán* a un ser de una interpretación de alto orden, o ir en busca de los tipos deliciosos de evidencia que *ascienden* a los seres a interpretaciones de un orden más alto (véase Bennett, 1976). Por ejemplo, nos encanta enterarnos de que los micos machos

[1] Podemos sondear los límites de la clase de equivalencia-estímulo para esta respuesta, sustituyendo para el leopardo "normal" estímulos tan diferentes como perros, hienas, leones, leopardos satisfechos, leopardos enjaulados, leopardos teñidos de verde, petardos, palas, motociclistas. Si estos tests independientes son tests de la *especificación de ansiedad* o del *significado* de oraciones de una palabra en el lenguaje de los micos, depende de si nuestros tests para los otros componentes de nuestra atribución de enésimo orden, los operadores intencionales anidados, resultan positivos.

que están solos, viajan de una banda a otra (y, por tanto, están fuera del alcance, hasta donde ellos creen, del oído de otros micos), buscarán *silenciosamente* refugio entre los árboles. Esto es suficiente para la hipótesis aguafiestas de los aullidos de ansiedad por los leopardos. (Ninguna hipótesis sucumbe tan fácilmente, por supuesto. Las modificaciones ad hoc pueden salvar cualquier hipótesis, y es un asunto fácil soñar con mecanismos interruptores de *"contexto"* simples para el aullido de ansiedad por el leopardo que reserven la hipótesis de cero orden para otro día.) En el otro extremo del espectro, el mero hecho de que los micos tengan aparentemente tan pocas cosas diferentes que sepan decir, ofrece pocas perspectivas para descubrir alguna utilidad teórica real para una hipótesis tan fantasiosa como nuestra candidata de cuarto orden. Es únicamente en contextos o sociedades en las que hay que excluir (o incluir) posibilidades tales como la ironía, la metáfora, el chismorreo y la ilustración (uso de palabras con una "segunda intención", como dirían los filósofos)[2] donde debemos valernos de interpretaciones de tan alto poder. Todavía no hay pruebas, pero habría que ser muy romántico para tener grandes esperanzas sobre esto. No obstante, hay anécdotas alentadoras.

Seyfarth informa (en el curso de la conversación) acerca de un incidente en el que una pandilla de micos estaba perdiendo terreno en una escaramuza territorial con otra pandilla. Uno de los monos del lado perdedor, momentáneamente fuera de la riña, pareció tener un idea brillante: emitió de repente una alarma para leopardos (sin que hubiera ningún leopardo) haciendo que *todos* los micos adoptaran el grito y se encaminaran hacia los árboles, provocando una tregua y recuperando el terreno que su lado había estado perdiendo. El sentido intuitivo que todos tenemos de que éste es *posiblemente* (excluyendo la interpretación aguafiestas) un incidente de gran inteligencia, se presta a un diagnóstico detallado en términos de los sistemas intencionales. Si este hecho no es sólo una feliz coincidencia, es verdaderamente tortuoso porque no se trata simplemente del mico lanzando un *imperativo* "métanse entre los árboles" con la esperanza de que *todos* los micos obedecieran, puesto que el mico (al ser racional, nuestra palanca predictiva) no debería *esperar* que una pandilla rival hiciera honor a *su* imperativo. De manera que la advertencia de leopardos es *considerada* por los micos informativa —un *aviso*, no una *orden*— y por tanto la credibilidad del emisor y no su autoridad es suficiente para explicar el efecto o el emisor del llamado es todavía más tortuoso: *quiere* que los rivales *piensen* que están *alcanzando a oír* una orden *destinada* (por supuesto) sólo para los suyos, etcétera. ¿Podría ser posible que un mico tuviera un sentido tan agudo de la situación? Estas alturas vertiginosas de sofisticación están implicadas estrictamente por la interpretación de orden más alto tomada con su inevitable presunción de racionalidad. Sólo de un ser capaz de apreciar estos aspectos se podría decir con exactitud que tiene esas creencias, deseos e intenciones.

Otra observación de los micos saca a relucir este papel de la asunción de la racionalidad aun más claramente. Cuando supe por primera vez que los métodos de Seyfarth implicaban ocultar altavoces en el matorral y tocar

[2] Véase Quine, 1960, págs. 48-49 acerca de los casos de segunda intención como "la ruina de la lingüística teorica".

alarmas grabadas, consideré el éxito del método como un dato seriamente degradante, puesto que si los monos fueran realmente griceanos en su refinamiento, cuando jugaran papeles ante su público deberían quedar perplejos, impasibles, de algún modo desgarrados por los avisos incorpóreos que no procedían de ningún emisor conocido. Si olvidaran este problema no serían griceanos. Tal como un comunicador auténtico normalmente controla a ratos al auditorio en busca de indicios de que está comprendiendo el significado de la comunicación, un público auténtico controla periódicamente al comunicador en busca de indicios de que el significado que está comprendiendo es el significado que se le está entregando.

Para mi deleite, sin embargo, supe por Seyfarth que se había tenido gran cuidado en el uso de los altavoces para impedir que este tipo de cosa ocurriera. Los micos pueden reconocer de inmediato las llamadas particulares de su pandilla. Así es como reconocen la llamada para leopardos de Sam *como* la de Sam y no la de Tom. Queriendo darle a las grabaciones la mejor oportunidad de "funcionar" los experimentadores tuvieron mucho cuidado en tocar el aviso de Sam sólo cuando Sam no estaba ni claramente a la vista ni con la boca cerrada u ocupado en otra cosa, ni cuando los otros "sabían" que estaba lejos. Sólo si el auditorio podía suponer que Sam estaba realmente presente y emitiendo el aviso (aunque oculto de sus miradas), sólo si el auditorio podía *creer* que el que hacía ruido entre los arbustos era Sam, tocarían los experimentadores el aviso de Sam. Si bien esta notable paciencia y cautela son dignas de aplauso como método escrupuloso, uno se pregunta si eran realmente necesarias. Si un programa de grabaciones "más desprolijo" produjera resultados igualmente "buenos", éste sería en sí mismo un dato *degradante* muy importante. Habría que intentar esa prueba; si los monos se desconciertan y quedan impasibles ante los llamados grabados en las circunstancias escrupulosamente mantenidas, la necesidad de esas circunstancias apoyaría con fuerza la necesidad de que Tom, digamos, sí cree que el que hace ruido entre los arbustos es Sam, que los micos no sólo son capaces de creer esas cosas sino que *deben* creerlas para que se produzca la reacción observada.

La asunción de racionalidad proporciona así una manera de tomar las distintas hipótesis en serio, lo suficientemente en serio como para verificarlas. Al principio esperamos que hayan con seguridad fundamentos para el veredicto de que los micos son creyentes sólo en alguna forma atenuada (por comparación con nosotros los creyentes humanos). La asunción de racionalidad nos ayuda a buscar hasta cierto punto, los indicios de la atenuación. Armamos condicionales tales como

14) Si x creyera que p y si x *fuera racional*, puesto que "p" implica "q" x creería (tendría que creer) que q.

Esto lleva a la ulterior atribución a x de la creencia que q,[3] la cual unida a alguna atribución plausible de deseo, lleva a una predicción de conducta

[3] "Siempre atesoraré el recuerdo visual de un filósofo muy enojado que estaba tratando de convencer a un auditorio de que 'si usted cree que *A* y cree que si *A* luego *B* y entonces usted *debe* creer que *B*.' Yo no sé verdaderamente si él tenía el poder moral para coercionar a nadie para que creyera que *B*, pero no poder cumplir realmente hace muy difícil usar la palabra 'creencia' y vale la pena gritar por eso" (Kahneman, inédito).

que se puede verificar por medio de la observación o el experimento.[4]

Una vez que se logra el don de utilizar la asunción de racionalidad para la eficacia, es fácil generar conductas reveladoras ulteriores a las que buscar en el estado natural o para provocarlas en experimentos. Por ejemplo, si algo tan sutil como un análisis de tercer o cuarto orden es correcto, debería ser posible entonces por el uso tortuoso (¡y moralmente dudoso!) de los altavoces ocultos crear un "muchacho que gritó lobo".[5] Si se elige un solo mico y "se lo hace aparecer" como el emisor de alarmas falsas, los demás, al ser racionales deberían empezar a perderle confianza, lo que debería manifestarse de distintas maneras. ¿Se podría crear un "vacío de credibilidad" para un mico? ¿El interés que tendría un resultado positivo así justificaría los resultados potencialmente desagradables (recuerde lo que ocurrió en la fábula)?

Cómo usar la evidencia anecdótica: el método Sherlock Holmes

Una de las trampas reconocidas de la etología cognitiva es el enojoso problema de la evidencia anecdótica. Por una parte, como buen científico, el etólogo sabe cuán engañosas y oficialmente inútiles son las anécdotas, y sin embargo, por otra parte, ¡a veces son tan reveladoras! El problema de los cánones de la evidencia científica es que excluyen virtualmente la descripción de todo lo que no sea el comportamiento muy repetido, muy observado y estereotípico de una especie y éste es exactamente el tipo de comportamiento que no revela absolutamente ninguna inteligencia especial. Todo este comportamiento se puede explicar, de manera más o menos verosímil, como los efectos de cierta combinación aburrida de "instinto" o tropismo y respuesta condicionada. Son los aspectos *innovadores* de la conducta, los hechos que no podrían ser explicados de manera verosímil en términos de condicionamiento, entrenamiento o hábitos anteriores, que hablan elocuentemente de inteligencia; pero si su originalidad e irrepetibilidad mismas los convierten en evidencia anecdótica y por lo tanto inadmisible, ¿cómo se puede desarrollar el tema cognitivo en favor de la inteligencia de la especie-objetivo?

Un problema exactamente así desesperó a Premack y Woodruff (1978), por ejemplo, en sus tentativas por demostrar que los chimpancés "tienen una teoría de la mente". Sus esfuerzos escrupulosos para obligar a sus chimpancés a una conducta repetible, no anecdótica que ellos creen que los chimpancés tienen engendra el frustrante efecto secundario de proporcionar historias de entrenamiento prolongado que los conductistas pueden señalar al desarrollar sus hipótesis condicionantes rivales como explicaciones putativas de la conducta observada.

Podremos ver la salida de este dilema si nos detenemos a preguntarnos

[4] La normalidad inadvertida de la presunción de racionalidad en cualquier atribución de creencias se revela señalando que 14) que asume explícitamente la racionalidad es virtualmente sinónimo de (juega el mismo papel que) el condicional que empieza: si *x* creyera realmente que *p*, entonces puesto que "*p*" implica "*q*"...

[5] Le debo esta sugerencia a Susan Carey en el transcurso de una conversación.

cómo fijamos nuestra *propia* intencionalidad de un orden más alto para satisfacción de todos, excepto los conductistas más doctrinarios. Podemos concederles a los conductistas que a cualquier breve intervalo único de la conducta humana se le puede dar una explicación relativamente verosímil y no obviamente degradante *ad hoc*, pero a medida que amontonamos anécdota sobre anécdota, novedad aparente sobre novedad aparente levantamos para cada persona conocida una biografía tal de inteligencia *aparente* que la afirmación de que *todo* no es más que una feliz coincidencia —o el resultado de un "entrenamiento" no descubierto hasta ahora— se convierte en la hipótesis más descabellada. Este agregado de detalles irrepetibles se puede instigar usando la actitud intencional para provocar circunstancias únicas que serán particularmente reveladoras. La actitud intencional es, en efecto, un motor generador o diseñador de circunstancias anecdóticas —artimañas, trampas y otras pruebas de tornasol intencionalistas— y predecir sus resultados.

Esta táctica tramposa ha sido celebrada en la literatura durante mucho tiempo. La idea es tan vieja como Ulises cuando probaba la lealtad de su porquerizo ocultando su identidad de éste y ofreciéndole tentaciones. Sherlock Holmes fue un maestro de experimentos intencionales más intrincados, de manera que llamaré a éste el *método Sherlock Holmes*, Cherniak (1981) atrae nuestra atención hacia un bonito vaso.

> En *Un escándalo en Bohemia* el rival de Sherlock Holmes ha escondido una fotografía muy importante en una habitación, y Holmes quiere descubrir dónde está. Holmes hace que Watson arroje una bomba de humo en la habitación y grite "fuego" mientras el adversario de Holmes está en el cuarto de al lado y Holmes observa. Luego, como era de esperar, el adversario entra corriendo en la habitación y quita la fotografía de donde estaba escondida. No cualquiera habría ideado un plan tan ingenioso para manejar la conducta de un rival; pero una vez que las condiciones están descritas parece muy fácil predecir las acciones del adversario (pág. 161).

En este ejemplo Holmes se entera simultáneamente de la colocación de la fotografía y confirma un perfil intencional bastante complicado de su adversario, Irene Adler, a quien se revela como *queriendo* la fotografía; que *cree* que está donde ella la va a buscar; que *cree* que la persona que gritó "fuego" *creyó* que había un incendio (observa que si ella creyera que quien gritó quería engañarla hubiera actuado de manera completamente distinta); *querer* recuperar la fotografía sin dejar que nadie *supiera* que ella estaba haciendo eso, etcétera.

Una variación sobre este tema es una táctica intencional amada por los escritores de temas de misterio: provocar el movimiento delator; todos los sospechosos están reunidos en la sala y el detective sabe (sólo él lo sabe) que el culpable (y sólo el culpable) *cree* que un gemelo de camisa incriminador está debajo de la pata de la mesa plegable. Por supuesto que el sospechoso no *quiere* que nadie más lo *crea*, o *descubra* el gemelo, y *cree* que a su debido tiempo será descubierto a menos que él tome una medida encubridora. El detective hace que haya un "corte de energía". Después de algunos segundos de oscuridad se encienden las luces y el culpable es, por supuesto, el sujeto

que está arrodillado debajo de la mesa plegable. ¿Qué otra cosa podría explicar de manera verosímil esta conducta tan nueva y extraña en un caballero tan distinguido?[6]

Se pueden planear estratagemas similares para verificar las distintas hipótesis acerca de las creencias y deseos de los micos y otros seres. Estas estratagemas tienen la virtud de provocar conductas novedosas pero que se pueden interpretar, de *generar anécdotas* en condiciones controladas (y, por tanto, científicamente admisibles). De este modo, el método Sherlock Holmes ofrece un aumento significativo del poder investigativo por encima de los métodos cognitivos. Esto aparece dramáticamente si comparamos la investigación real y meditada acerca de la comunicación en los micos con los esfuerzos del lingüista imaginario del campo conductista de Quine. Según Quine, un preliminar necesario para que el lingüista pueda hacer en verdad algún progreso es la aislación e identificación tentativa de palabras del idioma nativo (o actos del habla) en favor de "Sí" y "No", de manera que el lingüista pueda entrar en una ronda tediosa de 'preguntas y asentimientos', presentándoles o los naturales del lugar oraciones en su lengua en condiciones varias y tratando de encontrar patrones en sus respuestas por sí o por no (Quine, 1960, cap. 2). Los etólogos que estudian animales no pueden jugar a nada parecido al juego de preguntas y asentimientos de Quine, pero hay un vestigio evidente de esta estrategia minimalista de investigación en los pacientes estudios de "sustitución de estímulos" para las vocalizaciones animales, excluyendo, comúnmente, los experimentos más manipuladores (si bien menos intrusivos (véase nota 1). Mientras se sea decididamente conductista, sin embargo, no se entiende el valor de evidencia de una conducta tal como la del mico solitario metiéndose silenciosamente entre los árboles cuando se presenta un "estímulo por leopardos". Sin una cantidad considerable de una conducta reveladora así, ni una montaña de datos sobre lo que Quine llama el "significado estímulo" de las emisiones revelará que son actos de comunicación más que manifestaciones meramente audibles de sensibilidades

[6] Es un don especial del dramaturgo idear circunstancias en las que la conducta —verbal y de otra clase— habla en voz muy alta y claramente acerca de los perfiles intencionales ("motivación", creencias, malos entendidos, y así sucesivamente) de los personajes, pero a veces estas circunstancias se vuelven demasiado intrincadas para la comprensión inmediata. Un cambio muy leve en la circunstancia puede establecer una gran diferencia entre una conducta completamente inescrutable y una autorrevelación lúcida. El notorio "vete a un convento" del discurso de Hamlet a Ofelia es un caso clásico al respecto. El papel de Hamlet era totalmente desconcertante hasta que dimos con el hecho (oscurecido en las mínimas indicaciones escénicas de Shakespeare) de que mientras Hamlet está hablándole a Ofelia *cree* no sólo que Claudio y Polonio están escuchando detrás de las cortinas, *sino que creen que él no sospecha* de que lo están haciendo. Lo que hace que esta escena sea particularmente apta para nuestros objetivos es el hecho de que retrata un experimento intencional: Claudio y Polonio, usando a Ofelia como señuelo y puntal, están intentando provocar un comportamiento especialmente revelador de Hamlet para, a partir de ahí, descubrir cuáles son sus creencias e intenciones. Los frustra su incapacidad para diseñar el experimento lo bastante bien como para excluir del perfil intencional de Hamlet la creencia de que está siendo observado y el deseo de crear falsas creencias en sus observadores. Véase, por ejemplo, Dover Wilson, 1951. Una dificultad parecida puede confundir a los etólogos: "las observaciones breves de la conducta de la avoceta y la cigoñuela puede ser engañosa. Subestimando la aguda visión del pájaro los primeros naturalistas creyeron que su presencia había pasado inadvertida e interpretaron erróneamente la conducta distraída, como galanteo". (Sordahl, 1981, pág. 45).

determinadas. Por supuesto que Quine se da cuenta de esto, y presupone tácitamente que su traductor radical ya se ha convencido informalmente (sin duda utilizando el poderoso, pero cotidiano método Sherlock Holmes) de la naturaleza ricamente comunicativa de la conducta de los nativos del lugar.

Naturalmente, la fuerza del método Sherlock Holmes es de doble filo; no responder a las expectativas es, a menudo, un dato fuertemente degradante.[7] Woodruff y Premack (1979) han tratado de demostrar que los chimpancés de su laboratorio pueden ser *impostores* cabales. Téngase en cuenta a Sadie, una de los cuatro chimpancés utilizados en este experimento. A la vista de Sadie se coloca comida en una de dos cajas cerradas que ella no puede alcanzar. Luego entran un entrenador "cooperativo" o uno "competitivo" y Sadie ha aprendido que debe señalar una de las dos cajas si espera conseguir la comida. Si el entrenador cooperativo comparte la comida con Sadie. Sólo darle a Sadie la experiencia suficiente ante las circunstancias que le asegure la apreciación de estas contingencias implica sesiones de entrenamiento que le dan al conductista mucha experiencia para el "mero refuerzo" de su provecho. (Para que las identidades de los entrenadores se vuelvan suficientemente claras, hubo un estricto cumplimiento en el uso de determinada ropa y de rituales especiales; el entrenador competitivo siempre usó anteojos para el sol y una máscara de bandolero, por ejemplo. ¿Queda entonces la máscara fijada como un simple "estímulo provocador" de la conducta tramposa?).

Más aun, dejando de lado las redescripciones de los conductistas, ¿estará Sadie a la altura de las circunstancias y hará lo "correcto"? ¿Tratará de engañar al entrenador competitivo (y sólo a éste) *señalando la caja equivocada*? Sí, pero abundan las sospechas acerca de la interpretación.[8] ¿Cómo podríamos fortalecerla? Pues bien, si Sadie tiene en verdad la intención de engañar al entrenador, ella debe (al ser racional) empezar por la creencia de que el instructor todavía no sabe dónde está la comida. Supongamos, entonces, que introdujéramos a todos los chimpancés en un contexto completamente distinto, en cajas plásticas transparentes; ellos *deberían* llegar a *saber* que

[7] No quiero ser interpretado como *llegando a la conclusión* en este trabajo de que los micos o los chimpancés de laboratorio u otros animales no humanos *ya han demostrado* ser sistemas intencionales de un orden más alto. Una vez que se aplica el método Sherlock Holmes con imaginación y rigor, puede muy bien producir resultados que desilusionarán a los románticos. Estoy argumentando en favor de un método de formular preguntas empíricas y explicar el método demostrando lo que las respuestas *podrían ser* (y por qué). No doy esas respuestas anticipándome a la investigación.

[8] Es demasiado fácil detenernos demasiado pronto en nuestra interpretación intencional de un ser presumiblemente "inferior". Había una vez en una aldea un idiota que, cada vez que se le ofrecía elegir entre una moneda de 10 y una de 5, tomaba sin vacilar la de 5, ante las risas y las burlas de los espectadores. Un día alguien le preguntó cómo podía seguir siendo tan estúpido como para seguir eligiendo la moneda de cinco después de oír todas esas risas. Y el idiota replicó: "¿Usted cree que si yo alguna vez tomara la moneda de 10 me volverían a dar a elegir otra vez?".

Los rituales extrañamente desprovistos de motivaciones que acompañaron el entrenamiento de los chimpancés como fueron relatados en Woodruff y Premack (1979), podrían muy bien haber confundido a los chimpancés por razones similares. ¿Puede un chimpancé preguntarse por qué estos seres humanos no se limitan a comer la comida que está bajo su control? Si así fuera, semejante curiosidad podría trastornar las oportunidades de los chimpancés para entender la circunstancia en el sentido en que los investigadores lo estaban esperando. De lo contrario, este mismo límite en su comprensión de esos agentes y predicamentos, de algún modo socava la atribución de un estado tan sofisticado y de un orden tanto más alto como el deseo de engañar.

puesto que ellos —y cualquier otro— puede ver a través de ellas, cualquiera puede ver y, por lo tanto llegar a *saber* qué hay en ellas. Entonces, sobre la base de un nuevo test conductista y de una sola prueba, podemos introducir un día una caja plástica y una opaca colocar la comida en la caja plástica. Después entra el instructor competitivo y permite que Sadie lo vea mirar directamente la caja plástica. Si Sadie *sigue* señalando la caja opaca, revela, lamentablemente, que en realidad no tiene una comprensión de las ideas complejas implicadas en el engaño. Por supuesto que este experimento todavía está imperfectamente planeado. Por un lado, Sadie podría señalar la caja opaca por desesperación, al no ver ninguna opción mejor. Para mejorar el experimento habría que introducir una opción que le pareciera mejor a ella sólo si la primera opción no tuviera ninguna posibilidad, como en este caso. Más aun, ¿no debería Sadie estar confundida por la extraña conducta del instructor competitivo? ¿No tendría que molestarle que el instructor competitivo, al no encontrar nada de comida donde ella señala se siente "malhumorado" en un rincón en lugar de verificar la otra? ¿No debería asombrarla descubrir que su treta sigue dando resultado? *Debería* preguntarse: ¿Es posible que el instructor competitivo sea tan estúpido? Se necesitan otros experimentos mejor planeados son Sadie y otros seres. [9] No queriendo llenar el irritante estereotipo del filósofo como un contestador de preguntas de café empíricas, no obstante, sucumbiré a la tentación de formular algunas predicciones. Después de otros estudios surgirá que los micos (y los chimpancés y los delfines, y todos los animales superiores no humanos), exhiben síntomas mezclados y confusos de una intencionalidad de orden más alto. Pasarán algunos tests de alto orden y fracasarán en otros; en algunos aspectos se revelarán alertas ante las sutilezas de tercer orden desilusionándonos, sin embargo con su imposibilidad de entender algunos puntos de segundo orden aparentemente aún más simples. No se confirmará claramente ningún conjunto preciso, "riguroso" de hipótesis intencionales de ningún orden. La razón por la cual estoy dispuesto a formular esta predicción no es que yo crea tener alguna comprensión especial de los micos u otras especies sino sólo que he notado, como cualquiera puede hacerlo, que lo mismo es verdad acerca de nosotros los seres humanos. No somos ejemplares so problemáticos de sistemas intencionales de tercer, cuarto o quinto orden, tenemos la enorme ventaja de ser usuarios charlatantes del lenguaje, seres a los que se los puede hacer sentar a un escritorio y hacerles contestar largos cuestionarios y cosas así. Nuestra capacidad misma de participar en interacciones lingüísticas de este tipo distorsiona seriamente nuestro perfil como sistemas intencionales, produciendo ilusiones de mucha mayor definición en nuestros sistemas operativos de representación de las que en realidad tenemos. (*Brainstorms*, caps. 3 y 16; véase también el cap. 3 en este volumen.) Espero que los resultados de los esfuerzos en las interpretaciones intencionales de los monos, como los de

[9] Este comentario sobre los chimpancés de Premack surgió de una discusión en la conferencia de Dahlem sobre inteligencia animal con Sue Savage-Rumbaugh, cuyos chimpancés, Austin y Sherman exhiben una conducta aparentemente comunicativa (Savage-Rumbaugh, Rumbaugh y Boysen, 1978, que está pidiendo a gritos análisis y experimentación por la vía del método Sherlock Holmes.

las interpretaciones intencionales de niños pequeños estén plagadas de brechas y lugares nebulosos, inevitables en la interpretación de los sistemas que son, después de todo, sólo imperfectamente racionales (véanse capítulos 2 y 3).

Sin embargo los resultados serán valiosos a pesar de sus vacíos y sus vaguedades. ¿Cómo y por qué? El perfil o caracterización de un animal —o en lo que a eso respecta, de un sistema inanimado— según la actitud intencional, puede ser considerada como lo que los ingenieros llamarían un conjunto de especiales especificaciones de un artefacto con cierta *aptitud* para el proceso de información total. Un perfil del sistema intencional dice, aproximadamente, *qué información* debería recibir, usar, recordar y transmitir el sistema. Alude a las maneras en que las cosas del mundo circundante deberán estar representadas —pero sólo en los términos de diferencias deducidas o deducibles, de discriminaciones factibles— y en absoluto en los términos del mecanismo real para hacer este trabajo (véase Johnson, 1981 acerca de la "descripción de tareas"). Estas especificaciones intencionales fijan entonces una tarea de planeamiento para el próximo tipo de teórico, el diseñador del sistema de representación.[10] Esta división del trabajo ya es conocida en ciertos círculos de la Inteligencia Artificial (IA). Lo que yo he llamado actitud intencional es lo que Newell (1982) llama "el nivel de conocimiento". Y, curiosamente, los mismos defectos, brechas y lugares irracionales en el perfil intencional de un animal menos que idealmente racional, lejos de crearle problemas al diseñador del sistema, le señalan los atajos y los tapones provisorios en los que la Madre Naturaleza ha confiado para diseñar el sistema biológico; por tanto, ellos simplifican la tarea del diseñador del sistema.

Supongamos, por ejemplo, que adoptemos la actitud intencional hacia las abejas, y notemos, maravillados, que parecen *saber* que las abejas muertas son un problema de higiene en la colmena. Cuando una abeja muere, sus hermanas *reconocen* que ha muerto y al *creer* que las abejas muertas constituyen un riesgo sanitario y *desean*, muy racionalmente, evitar los riesgos sanitarios, *deciden* que deben retirar inmediatamente la abeja muerta. Por tanto, lo hacen. Ahora bien, si se confirmara esa historia intencional fantasiosa, el diseñador del sistema de las abejas se enfrentaría con un trabajo sumamente difícil. Felizmente para el diseñador (si bien lamentablemente para los que son románticos con respecto a las abejas), resulta que una explicación de un orden muy inferior alcanza: las abejas muertas segregan ácido oleico; el olor del ácido oleico pone en marcha la subrutina en las otras abejas de "retirarla". Coloque un toque de ácido oleico sobre una abeja viva y sana, y ésta será arrastrada, pataleando y chillando, fuera de la colmena (Gould y Gould, 1982; Wilson, Durlach y Roth, 1958).

Alguien que esté en inteligencia artificial, al enterarse de esto, podría muy bien decir: "¡Oh, cuán conocido! Sé *exactamente* cómo diseñar sistemas que se conduzcan así. Los atajos como ése son mi equipo indispensable. "El hecho es que hay una semejanza misteriosa entre muchos de los descubrimientos de los etólogos cognitivos que trabajan con animales inferiores y las

[10] En los términos que desarrollo en el capítulo 3, la teoría del sistema intencional especifica un mecanismo semántico que debe entonces ser ejecutado —imitado, aproximadamente— por un mecanismo sintáctico diseñado por el psicólogo cognitivo subpersonal.

proezas mezcladas con estupidez con las que uno se topa en los productos típicos de la *IA*. Por ejemplo, Roger Schank (1976) relata una historia acerca de un "bicho" en *TALESPIN* un programa de escritura de cuentos escrito por James Meehan en el laboratorio de Schank en Yale, que produjo el siguiente cuento: "La Hormiga Henry tenía mucha sed. Caminó hasta la orilla del río donde estaba sentado su buen amigo el Pájaro Bill. Henry resbaló y cayó al río. La gravedad se ahogó. ¿Por qué "se ahogó la gravedad?" (!) Porque el programa utilizó un atajo habitualmente fiable de tratar a la gravedad como un *agente* innombrado que siempre está tirando cosas, y puesto que la gravedad (al contrario de Henry en el cuento) no tenía ningún amigo (¡!), no había nadie que la salvara cuando en el río empujó a Henry hacia abajo.

Hace varios años, en *Why Not the Whole Iguana* (Dennett, 1978d), sugerí que la gente que está en *IA* podría progresar más si pasara del modelado de microcompetencias humanas (jugar al ajedrez, contestar preguntas sobre béisbol, escribir cuentos para niños muy pequeños, etc.), a las competencias totales de animales mucho más simples. Sugerí entonces que sería sensato que la gente de *IA inventara* seres imaginarios simples y les resolviera todo el problema de la mente. En este momento me siento tentado a pensar que la verdad es probablemente más fructífera y, al mismo tiempo, asombrosamente, más dúctil que la ficción. Sospecho que si algunas de las personas que trabajan con las abejas y las arañas unieran fuerzas con algunas de las personas de *IA*, sería un sociedad mutuamente enriquecedora.

Una perspectiva biológica más amplia de la actitud intencional

Es hora de hacer un inventario de este festejo alegre de la actitud intencional como estrategia de la etología cognitiva antes de pasar a ciertas sospechas y críticas latentes. He afirmado que la actitud intencional está bien adaptada como para describir de manera predictiva, provechosa e iluminadora, la proeza cognitiva de los seres en sus entornos, y que, más aun, permite perfectamente una división del trabajo en la ciencia cognitiva de la clase adecuada: los etólogos de campo, dadas tanto su preparación como los tipos de pruebas que se pueden deducir de sus métodos, no están en condiciones de formular —y menos aun de probar— hipótesis positivas acerca de los verdaderos mecanismos *representacionales* en los sistemas nerviosos de sus especies. Esa clase de diseño de *hardware* y *software* es la especialidad de algún otro.[11] Sin embargo, la actitud intencional proporciona exactamente la entrecara correcta entre las especialidades: una caracterización de "caja negra" de las aptitudes conductistas y cognitivas que se pueden observar en el terreno, pero escondidas en un lenguaje que reprime de manera muy pesada (ideal) el diseño del mecanismo a colocar en la caja negra.[12]

[11] Yo debería reconocer, sin embargo, que en el caso de insectos, arañas y otros seres *relativamente* simples hay algunos biólogos que han logrado llenar este vacío de manera brillante. Por supuesto, el vacío es mucho más estrecho en los no mamíferos.

[12] En "How to Study Human Consciousness Empirically: Or Nothing Comes to Mind" (Dennett, 1982b), se explica con mayor detalle cómo las descripciones puramente "semánticas" restringen las hipótesis acerca de los mecanismos sintácticos en la psicología cognitiva.

Este resultado aparentemente feliz se logra, sin embargo, por medio de la decisión dudosa de arrojar los escrúpulos conductistas por la borda y cometer actos de descripción mentalista, plenos de presunciones de racionalidad. Más aun; ¡al que da este paso le importan tan poco los detalles de la comprensión fisiológica como a cualquier dualista (estremecimiento)! ¿Esto puede ser legítimo? Creo que ayudará a contestar esta pregunta postergarla por un momento y considerar la adopción de la actitud intencional en el concepto más amplio de la biología.

Un fenómeno que ilustrará muy bien la conexión que quiero trazar es la "ostentación de distracción", la conducta bien conocida, encontrada en muchas especies bien separadas de pájaros que anidan en el suelo, de fingir tener un ala rota para alejar a un depredador que se acerca al nido, de sus habitantes indefensos (Simmons, 1952; Skutch, 1976). Este parece ser un *engaño* por parte del pájaro y, por supuesto, se lo llama así comúnmente. Su objetivo es *burlar* al depredador. Ahora bien, si la conducta es *verdaderamente* engañosa, si el pájaro es un verdadero impostor, debe tener una representación muy sofisticada de la situación. La razón de ser de ese engaño es sumamente complicada, y, adoptando la táctica expositora útil de Dawkins (1976), de inventar "soliloquios", podemos imaginarnos el soliloquio del pájaro:

> Soy un pájaro de nido bajo, cuyos polluelos no se pueden proteger del depredador que los descubra. Este depredador que se acerca seguramente los descubrirá pronto a menos que yo lo distraiga; lo podría distraer su *deseo* de atraparme y comerme, pero únicamente si pensara que hubiera una posibilidad *razonable* de atraparme (no es ningún estúpido); adoptaría justamente esa *creencia* si yo le *diera alguna prueba de que* ya no puedo volar. Eso lo puedo lograr fingiendo tener un ala rota, etcétera.

¡Y hablamos de elaboración! Es extremadamente improbable que cualquier "impostor" con plumas sea un sistema intencional de inteligencia. Un soliloquio más realista para cualquier pájaro seguiría probablemente la línea de: "Allí viene un depredador". De repente siento la tremenda urgencia de bailar esa tonta danza del ala rota. ¿Me pregunto por qué? (Sí, ya lo sé, sería locamente romántico suponer que un pájaro pudiera estar a la altura de un metanivel así, preguntándose acerca de su compulsión repentina.) Ahora bien, exactamente la pregunta de cuán sensible es el sistema cognitivo de control de un pájaro a las variables pertinentes del entorno es una pregunta empírica abierta y estudiable. Si los pájaros adoptan la ostentación de distracción aun cuando hay un candidato manifiestamente mejor para el foco de atención del depredador (otro pájaro realmente herido u otra presa probable, por ejemplo), la conducta será desenmascarada como perteneciente a un orden muy bajo (como la respuesta de las abejas al ácido oleico). Si por el contrario los pájaros —algunos pájaros— exhiben una sofisticación considerable en el uso de esta estratagema (distinguiendo entre distintos tipos depredadores o quizá revelando darse cuenta del hecho de que no se puede burlar al mismo depredador con el mismo truco una y otra vez), nuestra in-

terpretación de orden más alto de la conducta como auténticamente engañadora se fomentará y hasta se confirmará.

Pero supongamos que resulte que la interpretación aguafiestas estuviera más cerca de la verdad: el pájaro tiene una especie de tropismo mudo. ¿Descartaríamos por consiguiente el rótulo "engaño" para la conducta? Sí y no. Ya no le *acreditaríamos a ese pájaro* una razón principal de engaño, pero esa razón de ser no desaparecerá así nomás. Es demasiado evidente que la *raison d'être* de este comportamiento instintivo es su poder engañador. Esa es la razón por la cual se desarrolló. Si queremos saber por qué esta extraña danza llegó a ser provocativa precisamente en estas ocasiones, su poder para engañar a los depredadores tendrá que ser destilado de la miriada de otros datos, conocidos, desconocidos e inescrutables en la larga historia de la especie. ¿Pero quién apreciaba este poder, quién *reconoció* esta razón de ser, sino el pájaro o sus antepasados personales? ¿Quién si no la Madre Naturaleza misma? Es decir: nadie. La evolución por selección natural "eligió" este diseño por esta "razón".

¿Es imprudente hablar de este modo? Llamo a esto el problema de las *razones de ser indecisas*. A veces empezamos con la hipótesis de que le podemos asignar cierta razón de ser a (la "mente" de) un ser determinado pero luego lo pensamos mejor: el ser es demasiado estúpido como para albergarla. No descartamos necesariamente la razón de ser. Si no es por ninguna coincidencia que se produjo la conducta astuta, pasamos la razón de ser el individuo al genotipo en desarrollo. Esta táctica es evidente si pensamos en otros ejemplos de engaño no conductistas. Nadie ha supuesto nunca que a las polillas y mariposas que tienen manchas en las alas se les ocurrió la brillante idea de la pintura de camuflaje y la pusieron en práctica. Sin embargo, la razón de ser engañosa está allí lo mismo y afirmar que está allí es afirmar que hay un territorio en el cual es *predictiva* y, por tanto, explicativa. (Por una discusión en relación con esto véase Bennett, 1976, secciones 52, 53, 62.) Podemos no notar esto sólo debido a la obviedad de lo que podemos predecir. Por ejemplo, en una comunidad donde hay murciélagos pero no pájaros para los depredadores, no esperamos encontrar polillas con manchas (puesto que como cualquier impostor racional sabe cualquier juego de manos visual se desperdicia en los ciegos y miopes).

La transmisión de la razón de ser del individuo el genotipo es por supuesto un viejo truco. Durante un siglo hemos hablado en forma casual de cómo las especies "aprenden" a hacer cosas, "probando" distintas estrategias y, por supuesto, la práctica figurativa no ha quedado restringida a los rasgos cognitivos o conductistas. Las jirafas se estiraron los cuellos y los patos tuvieron la sabiduría de hacer crecer una membrana entre los dedos de sus patas. Todas son sólo maneras figuradas de hablar, por supuesto, en el mejor de los casos atajos expositorios meramente dramáticos, se podría pensar. Pero sorprendentemente estas maneras figuradas de hablar se pueden tomar a veces mucho más seriamente de lo que la gente lo hubiera creído posible. La aplicación de las ideas de la teoría del juego y la teoría de la decisión, —por ejemplo, el desarrollo que Maynard Smith (1972, 1974) hace de la idea de las *estrategias evolutivas estables*— dependía de tomar en serio el hecho de que los modelos de largo plazo en la evolución descritos de manera figurada en

términos intencionales guardaban un parecido suficiente con los modelos de las interacciones de plazos cortos entre los agentes (racionales) (humanos) como para garantizar la aplicación de los mismos cálculos descriptivos normativos a ellas. Los resultados han sido impresionantes.

Defensa del "Paradigma panglossiano"

La estrategia que une la teoría del sistema intencional con esta clase de exploración teórica de la teoría evolutiva es la adopción deliberada de *modelos óptimos*. Ambas tácticas son aspectos del *adaptacionismo*, el "programa basado en la confianza en el poder de la selección natural como agente optimizador" (Gould y Lewontin, 1979). Como observa Lewontin (1978b), "los argumentos óptimos se han vuelto sumamente populares en los últimos quince años, y representan actualmente el tipo de pensamiento dominante".

Gould se ha unido a Lewontin, su colega de Harvard, en la campaña de éste contra el adaptacionismo, y llaman al uso de los modelos óptimos por los evolucionistas "el paradigma panglosiano" en homenaje al Dr. Pangloss, la punzante caricatura de Voltaire hizo en *Candide* del filósofo Leibnitz, quien afirmaba que este es el mejor de todos los mundos posibles. El Dr. Pangloss podía racionalizar cualquier calamidad o deformidad —desde el terremoto de Lisboa a la enfermedad venérea— y demostrar sin ninguna duda que todo era para bien. En principio nada podía demostrar que éste no fuera el mejor de todos los mundos posibles.

El caso que apuntaba contra el pensamiento adaptacionista por Gould y Lewontin, ha sido muy mal interpretado aun por algunos de los que lo adoptaron, tal vez debido a la extraña desigualdad entre el ataque de Gould y Lewontin y la benignidad de sus conclusiones y recomendaciones explícitas. Amontonan desdén por los supuestos disparates del conjunto de la mente adaptacionista, lo que lleva a muchos a suponer que su conclusión es que el pensamiento adaptacionista debería ser totalmente evitado. En realidad, los críticos de una versión anterior de este trabajo que afirmaban que mi posición era una versión del adaptacionismo, atrajeron mi atención hacia el trabajo de aquellos. Según Gould y Lewontin el adaptacionismo "ha demostrado estar en bancarrota total". Pero cuando me volví hacia esta supuesta refutación de mis presunciones fundamentales descubrí que la recapitulación final de los autores encuentra un lugar legítimo en la biología del pensamiento adaptacionista. El suyo es un llamado en favor del "pluralismo", en realidad una queja contra lo que ven como una concentración exclusiva sobre el pensamiento adaptacionista, al precio de pasar por alto otros importantes enfoques del pensamiento biológico. Pero sin embargo los argumentos que preceden a esta conclusión suave y completamente razonable, parecen estar mal preparados para apoyarla, puesto que están claramente presentados como si fueran ataque a la integridad fundamental del pensamiento adaptacionista, más que como un apoyo para la recomendación de que en el futuro todos debemos tratar de ser adaptacionistas más cuidadosos y pluralistas.

Más aun, cuando estudié los argumentos más de cerca, me sentí atacado por un sentimiento de *déjà vu*. Estos argumentos no eran nuevos, sino más

bien una repetición de la prolongada y polémica campaña contra el "mentalismo" de B. F. Skinner. ¿Podría ser, me pregunté, que Gould y Lewontin hayan escrito el último capítulo del conservadurismo positivista de Harvard? ¿Podría ser que hayan recogido la antorcha a la que Skinner, en retirada, ha renunciado? Dudo de que Gould y Lewontin consideren al descubrimiento de su afinidad intelectual con Skinner con ecuanimidad, sin impurezas[13] y no intento sugerir en absoluto que el trabajo de Skinner sea la inspiración consciente para el de ellos, pero repasemos el alcance de su acuerdo.

Lewontin (1978b) nos dice que uno de los problemas principales del adaptacionismo es que es demasiado fácil: "Los argumentos óptimos prescinden de la tediosa necesidad de saber algo concreto acerca del fundamento genético de la evolución", señala cáusticamente; una imaginación sana es el único requisito para esta clase de "relato de cuentos" especulativo y la verosimilitud es, a menudo, el único criterio de esas historias (Gould y Lewontin, 1979, págs. 153-54).

Skinner (1974) nos dice que uno de los inconvenientes principales del mentalismo es "la manera [mentalista] en que las situaciones se inventan simplemente tan a menudo. Es demasiado fácil". Siempre se puede soñar con una "explicación" mentalista verosímil de cualquier conducta y si su primer candidato no resulta, siempre se lo puede descartar y encontrar otra historia. O, como Gould y Lewontin (1979, pág.153) dicen acerca del adaptacionismo, "Puesto que la amplitud de las historias adaptativas es tan grande como son productivas nuestras mentes, siempre se pueden postular historias nuevas. Y si no se dispone inmediatamente de una historia, siempre se puede alegar una ignorancia temporal y confiar en que se va a presentar".[14]

Gould y Lewontin objetan que las afirmaciones aprioristas no puedan ser falsificadas. Skinner afirma lo mismo de las interpretaciones mentalistas. Y ambos objetan todavía más que esta excesiva facilidad para planear historias *distraiga la atención* de los detalles difíciles y las partes más sustanciales que la ciencia debería buscar: Gould y Lewontin se quejan de que el pensamiento adaptacionista distrae al teórico de la búsqueda de la evidencia de la evolución no adaptativa por la vía de la corriente genética, "la compensación material", y otras variedades de "inercia filética" y las restricciones arquitectónicas. En el caso de Skinner el mentalismo distrae al psicólogo de

[13] A pesar de sus diferencias manifiestas, Lewontin y Skinner comparten en realidad una desconfianza profunda por la teorización cognitiva. Lewontin cierra su examen laudatorio de *The Mismeasure of Man* (1981) de Gould en el *New York Review of Books* (22 de octubre de 1981) con un rechazo categórico de la ciencia cognitiva, un veredicto tan amplio e indiscriminatorio como cualquiera de los ***obiter dicta*** de Skinner: "No es fácil, dado el modo analítico de la ciencia, reemplazar la mente que tiene un mecanismo de relojería con algo menos tonto. Poner al día la metáfora convirtiendo los relojes en ordenadores no nos han llevado a ninguna parte. El rechazo al por mayor del análisis en favor del holismo oscurantista fue peor. Aprisionados por nuestro cartesianismo, no sabemos cómo pensar acerca del pensamiento (pág. 16).

[14] Esta objeción le resulta muy conocida a E. O. Wilson, quien señala: "Paradójicamente, la mayor trampa en el razonamiento sociobiológico es la facilidad con que se la ejecuta. Mientras las ciencias físicas se ocupan de resultados precisos que son habitualmente difíciles de explicar, la sociobiología tiene resultados imprecisos que se pueden explicar demasiado fácilmente por medio de muchos esquemas diferentes" (1975, pág. 20). Véase también la discusión de éstos en Rosenberg, 1980.

buscar la evidencia de las historias de refuerzo. Como Skinner (1971) se queja, "el mundo de la mente se roba el espectáculo", pág. 12.

Ambas campañas utilizan tácticas similares. A Skinner le gustaba sacar a relucir los peores abusos del "mentalismo" para el escarnio, tal como las "explicaciones" psicoanalíticas (en términos de creencias, deseos, intenciones, temores, etc., inconscientes) de síndromes que resultan tener simples causas hormonales o mecánicas. Estos son casos de extensión exagerada, gratuita y negligente del reino de lo intencional. Gould y Lewontin dan como un mal ejemplo sacar conclusiones precipitadamente y de manera desprolija tal como lo hace un adaptacionista, Barash (1976), en su intento de explicar la agresión de los azulejos de montaña, la invención de una táctica contra el "adulterio", completa con su razón de ser donde se pasó por alto una explicación mucho más simple y directa (Gould y Lewontin, 1979, pág. 154). También "culpan al programa adaptacionista por su fracaso en distinguir la utilidad corriente de las razones de origen", una crítica que es exactamente paralela a la afirmación (que no he encontrado explícitamente en Skinner, aunque es bastante común) de que la interpretación mentalista a menudo confunde la racionalización *post hoc* con las "verdaderas razones" de un sujeto, que por supuesto deben reformularse, en los términos de una historia de refuerzo anterior.

Finalmente, está la reincidencia, las concesiones no reconocidas a los puntos de vista que se están atacando, comunes a ambas campañas. Notoriamente Skinner se valió de modismos mentalistas cuando se adaptaban a sus propósitos explicativos, pero disculpó esta práctica como si fuera taquigrafía, o como palabras fáciles para beneficio de los legos, sin reconocer jamás cuánto tendría que dejar de decir si abjurara del todo de la discusión mentalista. Gould y Lewontin son mucho más sutiles: ellos adoptan el "pluralismo", después de todo, y ambos son muy claros acerca de la utilidad y probidad —y hasta la necesidad— de *algunas* explicaciones y formulaciones adaptacionistas.[15] Cualquiera que las lea como si pidieran la extirpación de raíz del adaptacionismo los lee seriamente mal, aunque declinan decir cómo reconocer una buena pizca de adaptacionismo de las pizcas que ellos deploran. Esta es indudablemente una profunda diferencia con Skinner, el enemigo implacable del "mentalismo". Pero, sin embargo, me parece que no reconocen completamente su propia confianza en el pensamiento adaptacionista, o por cierto, su posición central en la teoría evolucionista.

Esto aparece muy claramente en el (merecidamente) popular libro de ensayos de Gould *Ever since Darwin* (1977). En "Darwin's Untimely Burial" Gould demuestra hábilmente cómo salvar la teoría darwinista de ese viejo espantajo de su reducción a una tautología, por la vía de un concepto vacío de idoneidad: "ciertos rasgos morfológicos, fisiológicos y conductistas debe-

[15] Lewontin, por ejemplo, cita su propio primer trabajo adaptacionista, *Evolution and the Theory of Games* (1961) en su crítica reciente a la sociobiología, *Sociobiology as an Adaptationist Program* (1979). Y en su artículo del *Scientific American*, "Adaptation" llega a esta conclusión: "Abandonar completamente la noción de adaptación, para observar simplemente el cambio histórico y describir sus mecanismos totalmente en los términos de los distintos triunfos reproductivos de diferentes tipos, sin ninguna explicación funcional, sería como tirar al bebé junto con el agua del baño" (1978a, pág. 230).

rían ser superiores *a priori* como diseños para vivir en ambientes nuevos. Estos rasgos confieren aptitud de acuerdo con criterio de buen diseño de un ingeniero, no de acuerdo con el hecho empírico de su supervivencia y difusión" (1977, pág. 42). [16] De manera que podemos mirar los diseños del modo en que lo hacen los ingenieros y evaluarlos como mejores o peores sobre la base de cierto conjunto de presunciones acerca de las condiciones y necesidades o de los objetivos. Pero eso es adaptacionismo. ¿Es pandossiano? ¿Compromete a Gould con el punto de vista de que los diseños elegidos siempre producirán el mejor de todos los mundos posibles? El repudio habitual en la literatura es que la Madre Naturaleza no es optimizadora sino "satisfactoria" (Simon, 1957), un conciliador en favor *de lo mejor* que se tiene a mano, lo bastante bueno, y no un rigorista de lo *óptimo*. Y si bien siempre vale la pena destacar esto, deberíamos recordar la vieja broma panglossiana: el optimista dice que éste es el mejor de los mundos posibles; el pesimista suspira y asiente.

La broma revela vívidamente la existencia inevitable de un trueque entre las restricciones y lo óptimo. Lo que parece estar muy lejos de lo óptimo en un conjunto de restricciones *puede* verse como óptimo en un conjunto más grande. Los torpes arreglos provisionales con los cuales el velero desarbolado vuelve renqueando a puerto pueden parecer un diseño mediocre para un velero hasta que reflexionamos que dadas las condiciones y los materiales disponible, lo que estamos viendo puede ser exactamente el mejor diseño posible. Por supuesto que puede no serlo. Tal vez los marineros no sabían hacerlo mejor, o se desconcertaron y se conformaron con arreglos, claramente inferiores. Pero, ¿qué pasa si aceptamos esa ignorancia de los marineros como condición límite? "Dada su ignorancia de las bondades de la aerodinámica, es probable que ésta fuera la mejor solución que *ellos* pudieron ver." ¿Cuándo dejamos —o debemos dejar— de agregar condiciones? En principio no hay un límite que yo pueda ver, pero no creo que éste sea un retroceso *perverso*, puesto que estabiliza y se detiene normalmente después de algunos movimientos, y dure cuanto dure, los descubrimientos que provoca son potencialmente iluminadores.

No *suena* panglossiano recordarnos, como Gould a menudo lo hace, que la vieja y pobre Madre Naturaleza se las arregla explotando oportunísticamente y con falta de perspicacia cualquier cosa que tenga a mano, hasta que agregamos: no es perfecta, pero *hace lo mejor que puede*. Satisfacerse puede demostrar a menudo ser la estrategia *óptima* cuando "los costes de la investigación" se agregan como una restricción (véase Nozick, 1981, pág. 300 para una discusión). Gould y Lewontin tienen razón en sospechar que hay un mecanismo tautológico en las bambalinas del teatro adaptacionista, siempre listo para hacer girar un nuevo conjunto de restricciones que salvará la visión panglossiana, pero creo que están comprometidos a actuar en el mismo escenario, por más cautelosamente que controlen sus papeles.

Skinner está también igualmente en lo cierto cuando insiste en que *en principio* las explicaciones mentalistas no se pueden falsificar: su estructura lógica *siempre* permite un repaso *ad lib* para preservar la racionalidad. De

[16] Para una discusión rigurosa de cómo definir la aptitud de manera de evadir la tautología, véase Rosenberg, 1980, págs. 164-75.

este modo si predigo que Joe vendrá a clase hoy porque quiere obtener una buena nota, y cree que se presentará material importante; y Joe no se presenta, no hay nada más fácil que decidir que *debe de* haber tenido algún compromiso más urgente, o no debe de haber sabido la fecha de hoy o simplemente que se debe de haber olvidado o... hay otras mil hipótesis más disponibles sin demora. Por supuesto que lo puede haber atropellado un camión, en cuyo caso más interpretaciones intencionales alternativas no son mucho más que versiones. Los peligros señalados por Skinner y por Gould y Lewontin, son reales. Los adaptacionistas, como los mentalistas, corren en realidad el riesgo de levantar sus edificios teóricos prácticamente de la nada, poniéndose en ridículo cuando esos castillos de naipes se derrumban, tal como de vez en cuando ocurre. Este es el riesgo que siempre se corre cuando se adopta la actitud intencional o la adaptacionista, pero puede ser prudente aceptar el riesgo puesto que el rédito es con frecuencia tan alto, y la tarea que el teórico más cauteloso y sobrio tiene que enfrentar es tan extraordinariamente difícil.

El adaptacionismo y el mentalismo (teoría del sistema intencional) no son *teorías* en un sentido tradicional. Son actitudes o estrategias que sirven para ordenar datos, explicar interrelaciones y generar preguntas para formularle a la Naturaleza. Si fueran teorías según el molde "clásico", la objeción de que son imploradores de preguntas o irrefutables sería fatal, pero formular esta objeción es interpretar mal la lectura de su objetivo. En un artículo muy perspicaz Beatty (1980) cita a los adaptacionistas Oster y Wilson (1978): "'El curso prudente es considerar los modelos de optimización como guías provisionales para la investigación empírica futura y no como la clave para leyes más profundas de la naturaleza'" (pág. 312). Se puede decir exactamente lo mismo acerca de la estrategia de adoptar la actitud intencional en la etología cognitiva.

La crítica de una vacuidad siempre amenazadora, levantada tanto contra el adaptacionismo como el mentalismo sería verdaderamente reveladora si en realidad siempre, o hasta con mucha frecuencia, nos aprovecháramos de las languidez disponible en principio. Si estuviéramos reconsiderando, *post hoc*, nuestros perfiles intencionales de las personas cuando no hicieran lo que esperábamos, la práctica revelaría ser una farsa, pero entonces, si ése fuera el caso, la práctica se hubiera extinguido hace mucho tiempo. Del mismo modo, si los adaptacionistas se vieran obligados a reconsiderar siempre (o con mucha frecuencia) sus listas de restricciones *post hoc* para preservar su panglossianismo, el adaptacionismo sería una estrategia muy poco atractiva para la ciencia. Pero la realidad acerca de ambas tácticas es que, en pocas palabras, *funciona bien*. No siempre, pero con gratificadora frecuencia. Somos realmente muy hábiles para escoger las restricciones correctas, las atribuciones correctas de creencia y deseo. La prueba del esfuerzo propio para la afirmación de que hemos localizado verdaderamente todas las restricciones importantes en relación con las cuales debería calcularse un diseño óptimo, es que hagamos ese cálculo optimizador y resulta ser predictivo en el mundo real. Se afirma haber localizado todas las restricciones auténticamente importantes sobre la base de que

1) el diseño óptimo dadas esas restricciones es *A*
2) la Madre Naturaleza optimiza
3) *A* es el diseño observado (es decir, aparente).

Aquí se da por sentado a Pangloss para inferir la terminación de la lista de restricciones. ¿Qué otro argumento se podría usar jamás para convencernos a nosotros mismos de que habíamos localizado y apreciado todas las consideraciones pertinentes en la historia evolutiva de alguna característica? Como dice Dawkins (1980, pág. 358), una teoría adaptacionista tal como la teoría de la estrategia evolutivamente estable de Maynard Smith:

> no intenta ser en general una hipótesis verificable que puede ser verdadera o falsa; la evidencia empírica lo decidirá. Es una herramienta que podemos utilizar para descubrir las presiones selectivas que se ejercen sobre la conducta animal. Como ha dicho Maynard Smith (1978) sobre la teoría de la optimización en general: "*No* estamos verificando la proposición general que la naturaleza optimiza, sino las hipótesis específicas acerca de las restricciones, los criterios de optimización y la herencia. Habitualmente verificamos si hemos identificado correctamente las fuerzas selectivas responsables".

Los peligros de la ceguera en el pensamiento adaptacionista, señalados tan vívidamente por Gould y Lewontin se reflejan como en un espejo en cualquier enfoque que rehúye la curiosidad adaptacionista. Dobzhansky (1956) dice, muy en el espíritu de Gould y Lewontin: "La utilidad de una característica debe demostrarse, no se la puede dar por sentada". Pero, como observa Cain (1964): "Del mismo modo no puede darse por sentada su inutilidad y la evidencia indirecta acerca de su probabilidad de ser elegida y realmente adaptativa, no puede ser pasada por alto... Allí donde se han llevado a cabo investigaciones, las características triviales han demostrado ser de importancia adaptiva por derecho propio." Cain compara solapadamente la actitud de Dobzhansky con la curiosidad de Robert Hooke por las antenas de los insectos en *Micrographia* (1965):

> No puedo imaginarme bien cuál sería el uso de esta clase de cuerpos con cuernos y penachos, a menos que sirvan para olfatear u oír, aunque parece muy difícil describir cómo se adaptan a cualquiera de las dos cosas; están en casi todas las distintas clases de moscas de las más variadas formas, aunque por cierto son una parte esencial de la cabeza, y tienen una misión notable que la Naturaleza les asignó, puesto que se las ha de encontrar en una u otra forma en todos los insectos.

"Aparentemente", infiere Cain, "la actitud correcta hacia los órganos enigmáticos pero que aparecen mucho, ya era entendida hace tanto como a mediados del siglo XVII al menos en Inglaterra" (1964, pág. 50).

Finalmente, me gustaría llamar la atención hacia un aspecto importante que Gould señala acerca del *sentido* de la biología, la pregunta fundamental que los evolucionistas deberían formular de manera persistente. Esto ocurre en su explicación aprobatoria del brillante análisis adaptacionista (Lloyd y Dibas, 1966) del hecho curioso de que los ciclos reproductivos de la cigarra tienen una duración de números primos: trece años por ejemplo, y diecisiete años: "Como evolucionistas, buscamos respuestas a la pregunta:

por qué. ¿Por qué, en especial, debería evolucionar una sincronicidad tan llamativa, y por qué tendría que ser tan largo el período entre los períodos de reproducción sexual?" (Gould, 1977, pág. 99). Como lo demuestra su propia explicación, *todavía* no se ha contestado la pregunta "por qué" planteada cuando uno ha emprendido sobriamente el largo viaje (en realidad extremadamente inaccesible) de la historia de la mutación, depredación, reproducción, selección, sin ningún lustre adaptacionista. Sin el lustre adaptacionista, no *sabremos por qué.*[17]

El contraste entre los dos tipos de respuestas, la respuesta histórico-arquitectónica escrupulosamente no-adaptacionista que Gould y Lewontin *parecen* estar defendiendo y la respuesta adaptacionista francamente panglossiana que uno también puede tratar de dar, se capta vívidamente en una analogía final de la guerra skinneriana contra el mentalismo. Una vez me encontré yo mismo en un debate público con uno de los discípulos más devotos de Skinner, y en un punto respondí a uno de sus más atrozmente inverosímiles skinnerismos con la pregunta: "¿Por qué dice *eso*?". Su respuesta instantánea y elogiosamente devota fue: "Porque fui reforzado para decir eso en el pasado". Mi pregunta "por qué" solicitaba una justificación, una razón de ser, no meramente un relato de origen histórico. Es posible, por supuesto, que cualquier pregunta especial "por qué" así reciba la respuesta: por "ninguna *razón*" en absoluto. Simplemente ocurrió que algo me hizo decir eso, pero la verosimilitud de semejante respuesta cae casi hasta el cero a medida que la complejidad y la significación aparente de la elocución sube. Y cuando se encuentra una razón de ser tolerable para un acto así es un error —una aplicación equivocada anacrónica del positivismo— insistir en que la "verdadera razón" para el acto *debe* ser manifestada en términos que no hagan ninguna alusión a esta razón de ser. Una explicación puramente causal del acto al nivel microfísico, digamos, *no está en competencia* con la explicación de la razón de ser. Hoy por hoy, esto lo entienden generalmente los psicólogos y filósofos posconductistas pero el punto contrario todavía no ha sido tan bien recibido entre los biólogos, a juzgar por el siguiente trozo aparecido en *Science*, al informar acerca de la famosa conferencia de 1980 en Chicago sobre la macroevolución:

> ¿Por qué tienen cuatro patas la mayoría de los vertebrados terrestres? La respuesta aparentemente obvia es que esta disposición es el diseño óptimo. No obstante, esta respuesta pasaría por alto el hecho de que los peces que fueron antecesores de los animales terrestres también tienen cuatro extremidades o aletas. Las cuatro extremidades pueden ser muy adecuadas para la locomoción en tierra firme, pero *la verdadera razón* [el subrayado es mío] por la cual los animales terrestres tienen esta disposición, es porque sus antecesores evolutivos tenían el mismo modelo (Lewin, 1980, pág. 886).

[17] Boden (1981) anticipa las afirmaciones en favor de la "actitud cognitiva" (en esencia, lo que he llamado la actitud intencional) en una localización biológica diferente: la microestructura de la genética, la ubicación de "reconocimiento" de las enzimas, la embriología y la morfogénesis. Como él dice, la actitud cognitiva "puede animar a los biólogos a formular preguntas empíricamente provechosas, preguntas que un enfoque puramente físico-químico podría tender a dejar sin formular" (pág. 89).

Cuando los biólogos formulan la pregunta "por qué" de los evolucionistas, están buscando, como los mentalistas, la razón de ser que explique por qué se eligió determinada característica. Cuando más compleja y aparentemente plena de significado es la característica, menor es la probabilidad de que no haya ninguna razón de ser que la sustente; y si bien los hechos históricos y arquitectónicos de la genealogía pueden en muchos casos vislumbrarse como los hechos más salientes o importantes a destapar, la verdad de esa historia no adaptacionista no exige la mentira de todas las historias adaptacionistas de la misma característica. La respuesta *completa* a la pregunta de los evolucionistas casi siempre aludirá al *mejor* diseño en una forma por lo menos mínima.

¿Es éste el mejor de todos los mundos posibles? No deberíamos ni siquiera tratar de contestar esa pregunta, pero adoptar la presunción de Pangloss, y en particular la presunción panglossiana de racionalidad en nuestros compañeros cognizadores, puede ser una estrategia inmensamente provechosa en la ciencia, sólo si podemos abstenernos de convertirla en un dogma.

Reflexiones: Interpretando a los monos, los teóricos y los genes

Cuatro años de discusión y estudio, tanto en el terreno cuanto en la biblioteca, han ampliado mucho mi perspectiva acerca de los puntos planteados en el ensayo anterior. En estas reflexiones, después de atar algunos cabos sueltos acerca del encuadre del ensayo y de cómo fue recibido, volveré a contar lo que aprendí de primera mano acerca de los problemas prácticos de adoptar la actitud intencional con los micos de Kenia, y luego repasaré el estado de la polémica que rodea al adaptacionismo y la actitud intencional, y por fin extraeré algunas implicaciones ulteriores acerca de su relación íntima.

Los ancestros y la progenie

En "Conditions of Personhood" (1976) manifesté que nuestras intenciones de un orden más alto señalaban una diferencia crítica entre nosotros y otras bestias y especulé sobre la pregunta empírica de exactamente cómo se podría confirmar la presencia de esas intenciones en los animales no humanos. Yo había discutido el tema con David Premack en 1975 y después hice comentarios (1978b) acerca de su artículo en *Behavioral and Brain Sciences* (*BBS*): "¿El chimpancé tiene una teoría de la mente?" (Premack y Woodruff, 1978), pero la Conferencia Dahlem sobre "Mente animal- Mente humana" en marzo de 1981 (Griffin, 1982) fue mi primera presentación de largo alcance en relación con los problemas y posibilidades de la interpretación en etología, psicología animal y ecología. En esa conferencia fui abochornado e invitado —o quizá desafiado— a demostrar cómo la adopción consciente de la actitud

intencional podría de verdad ayudar a los científicos a planear experimentos o interpretar sus datos. Me alegró descubrir que mis experimentos improvisados, al aplicar la actitud intencional a sus problemas de investigación, generaban en realidad algunas hipótesis innovadoras verificables, planes para experimentos futuros y métodos para desarrollar interpretaciones. Los participantes en la conferencia me exhortaron a escribir una introducción a la actitud intencional que fuera accesible para los no filósofos, y el capítulo anterior, vuelto a imprimir de *BBS*, es el resultado.

Como todos los "artículos meta" de ese periódico, éste fue acompañado por una amplia gama de comentarios y una réplica del autor [véase también *Continuing Commentary*, *BBS* 8 (1985b) págs. [758-66]. Las varias docenas de comentaristas incluyen filósofos, etólogos, psicólogos y teóricos evolucionistas, y sus críticas sondean los puntos fuertes y débiles de mi punto de vista desde muchos ángulos, un recurso valioso para cualquiera que desee proseguir más cuidadosamente con los temas aquí planteados. Recomiendo en especial los comentarios de Bennett, Menzel y Lewontin, los tres críticos más vehementes —desde tres lugares de privilegio totalmente diferentes, pero también recomiendo mis réplicas a ellos. Los comentarios de Skinner y Rachlin dan una perspectiva útil de los conductistas, mientras que Dawkins, Eldredge, Ghiselin y Maynard Smith iluminan los debates entre los teóricos evolucionistas.

Lo ocurrido desde que la obra apareció en 1983 hace que sus puntos fuertes y débiles resalten más claramente. Cheney y Seyfarth extendieron sus análisis del sistema de comunicación de los micos en una serie de artículos (1982,1985) y más recientemente han tenido éxito en encontrar una buena variación del experimento sobre "el muchacho que gritó ¡Lobo!". Da la casualidad de que los micos tienen dos gritos acústicamente muy diferentes con (aparentemente) el mismo significado; se acerca una pandilla rival de micos. No es sorprendente que despúes de escuchar varias veces la misma grabación del grito individual de un mico se cree hábito entre los oyentes, medido como una disminución gradual de las reacciones de vigilancia y otras por el estilo, pero puesto que hay dos gritos "sinónimos", se ha probado que es posible conseguir pruebas de la habituación no al *sonido* sino al *significado* del grito, y más aun, que la habituación está en relación con el emisor. Así es como un determinado mico puede perder credibilidad en el grupo con respecto a un tema determinado, gracias a estar "incriminado" por los investigadores (Seyfarth y Cheney), comunicación personal.

Mientras tanto, Ristau ha producido una serie de estudios experimentales sobre las exhibiciones de distracción de las avefrías (chorlitos) (Ristau, inédito, de próxima aparición, y Marley han investigado la sensibilidad de los pollos domésticos hacia un "auditorio" adecuado (Marler y otros, 1986a, 1986b). Byrne y Whiten (de próxima aparición) examinan una amplia variedad de estudios experimentales con y observación de, primates. Entre otras reacciones de los etólogos a mis propuestas, he encontrado especialmente útiles a Heyes (de próxima aparición) y McFarland (1984). Los psicólogos Wimmer y Perner (1983) han utilizado una variación del experimento que propuse para la verificación de creencias de orden más alto en Sarah, la chimpancé de Premack, para demostrar el notable comienzo de las creencias

acerca de las creencias en los niños pequeños. Y Premack ha llevado a cabo una gran cantidad de experimentos más que arrojan luz sobre el alcance y los límites de la comprensión que el chimpancé tiene de otros seres y un sistema de signos artificiales (Premack, 1986).

Otros etólogos y psicólogos han utilizado los métodos que recomendé con más entusiasmo que cuidado, lamento decirlo. Entiendo que unos pocos han confundido mi defensa del método Sherlock Holmes para "crear" (y "controlar") anécdotas ¡con una defensa al por mayor de anécdotas obtenidas casualmente, como pruebas! De manera que yo tendría que reiterar y recalcar en qué estaba haciendo hincapié: una muestra de conducta única y singular no sirve como *prueba* para una atribución de estado intencional (por más valiosa que pueda ser para el investigador como un indicio para experimentos futuros), a menos que se pueda *demostrar* que es una conducta improbable en otras circunstancias, sólo provocada por las condiciones que causarían, en un agente racional, creencia y deseos que convertirían en racional la conducta improbable. *Demostrar esto exige siempre el funcionamiento de experimentos controlados.* El método que yo alababa no era un sustituto de la experimentación, sino una manera de ver qué experimentos hacía falta hacer.

Otro fallo en mi tentativa de impartir refinamientos filosóficos a los estudiantes del comportamiento animal ha surgido de mis continuas discusiones con los etólogos: de una manera general persisten en juntar la noción filosófica (la noción de Brentano, el concepto de *acerquidad*, en una palabra) y la noción de intencionalidad más o menos cotidiana: la capacidad de ejecutar acciones intencionales o de armar intenciones para actuar. Hay pocos sermones más agotadores que los del filósofo tratando de reformar los hábitos lingüísticos de otros, y me he sentido penosamente tentado de abandonar el asunto, especialmene puesto que cuanto más rigurosamente se estudie la pregunta empírica de sí o cuándo un organismo es capaz de fraguar una intención de actuar más se combina, en la forma en que reúne y trata los datos, con la pregunta empírica acerca de si o cuándo un organismo tiene estados ("mentales") que exigen una caracterización por la vía de los modismos intencionales. Pero todavía pienso que es importante mantener la distinción entre estas dos maneras de plantear preguntas. En particular, de otro modo es demasiado fácil creer que se han eliminado las apelaciones a la intencionalidad cuando se canjea la discusión acerca de las intenciones, esperanzas y expectativas por la discusión aparentemente más científica sobre el almacenaje de información y estructuras meta.

En *Content and Consciousness* (1969) propuse reforzar la conciencia intensificada de esta distinción capitalizando siempre el término "intencionalidad" de los filósofos. Persistí en esta política en "Sistemas intencionales" (1971) pero no pude conseguir que el coche se moviera. Entonces, cuando volví a imprimir este artículo en *Brainstorms* (1978a) abandoné de mala gana mi idiosincrasia ortográfica solitaria. Desde entonces, Searle (1983) ha reinventado el esquema de capitalización, pero puesto que sus puntos de vista acerca de la intencionalidad no son en absoluto los míos, he decidido permitirle que se guarde la primera persona. Lo que Searle llama intencionalidad es algo en lo que no creo para nada. [Véanse capítulos 8 y 9, y mi comen-

tario sobre Searle en "The Milk of Human Intentionality" (1980b). Para un registro en una enciclopedia sobre el concepto de intencionalidad del filósofo, véanse Dennett y Haugeland (1987) en The Oxford Companion to the Mind.]

Abandono de la actitud teórica

En junio de 1983 pasé por una breve introducción al trabajo en el terreno etológico cuando observé a Seyfarth y Cheney estudiando a los micos en Kenia. (Esto está descrito en detalle en Dennett, de próxima aparición b y c, del cual está extractado el resto de esta sección.)

Una vez que llegué al terreno y vi de primera mano algunos de los obstáculos para llevar a cabo los tipos de experimentos que yo había recomendado, encontré algunas buenas noticias y otras malas. La mala noticia era que el método Sherlock Holmes en su aspecto clásico, tiene una aplicación muy limitada a los micos, y, por extrapolación, a otros animales "inferiores". La buena nueva era que adoptando la actitud intencional se pueden generar algunas hipótesis verosímiles e indirectamente verificables acerca de por qué esto debería ser así y de este modo aprender algo importante acerca de las presiones de la selección que han dado forma probablemente a los sistemas de comunicación de los micos.

Una vocalización que Seyfarth y Cheney estudiaban durante mi visita había sido llamada el gruñido MIO, de *Moving Into the Open* [saliendo al exterior]. Poco antes de que un mono que está oculto entre los arbustos salga al descampado a menudo emite un gruñido MIO. Otros monos que están entre los arbustos lo repiten con frecuencia. El análisis espectrográfico (todavía) no ha revelado una señal clara de diferencia entre el gruñido inicial y esta respuesta. Si ese eco no se produce, el gruñidor original a menudo se quedará entre los arbustos por cinco o diez minutos y luego repetirá el MIO. Cuando éste es repetido como un eco por uno o más monos, el gruñidor original se moverá cautelosamente hacia el campo raso.

¿Pero qué significa el gruñido MIO? Hicimos una lista de las traducciones posibles para ver cuáles podíamos eliminar o apoyar sobre la base de la evidencia de la que ya disponíamos. Comencé con la que parecía ser la posibilidad más directa y evidente.

"Salgo."

"Te interpreto. Estás saliendo."

Pero, ¿para qué serviría decir esto? Los micos son en realidad un grupo taciturno —se quedan callados la mayor parte del tiempo— y no se inclinan a nada que se parezca a saludar por medio de comentarios obvios. ¿Podría ser un pedido de permiso para salir?

"Por favor, ¿puedo salir?"

"Sí, te doy permiso para salir."

Esta hipótesis podría ser puesta fuera de combate si los micos de rango más alto emitieran alguna vez el MIO en presencia de sus subordinados. En realidad, los micos de más alto rango tienden efectivamente a salir primero al descampado, de manera que no parece que el MIO sea una solicitud de permiso. ¿Podría ser una orden entonces?

"¡Síganme!"

"¡A la orden, Capitán!"

No muy verosímil, pensó Cheney. "¿Por qué malgastar palabras con una orden así cuando parecería *sobreentendido* en la sociedad de los micos que los animales de bajo rango siguen la guía de sus superiores? Por ejemplo, se pensaría que debería haber una vocalización que significara 'Permiso' a ser dicha por un mono al acercarse a un superior con la esperanza de atenderlo. Y se esperaría que hubieran dos respuestas: 'Tiene permiso' y 'No tiene permiso' pero no hay indicios de ninguna vocalización así. Aparentemente estos intercambios no serían lo bastante útiles como para que valiera la pena el esfuerzo. Hay gestos y expresiones faciales que pueden cumplir este propósito, pero ninguna señal audible." Quizá, meditó Cheney el gruñido MIO servía simplemente para reconocer y compartir el miedo.

"Estoy realmente muy asustado."

"Sí. Yo también."

Otra posibilidad interesante sería que el gruñido ayudara a coordinar los movimientos del grupo.

"¿Están listos para que yo salga?"

"Estamos listos para cuando tú lo estés."

El mono que produce el eco puede ser el siguiente en salir. O quizá, mejor aun.

"¿Sin moros en la costa?"

"Sin moros en la costa. Te cubrimos."

La conducta observada hasta ahora es compatible con esta lectura que le daría al gruñido MIO un objetivo fuerte, orientando a los monos hacia una tarea de vigilancia cooperativa. Los monos que contestan, realmente vigilan al que sale y miran en las direcciones correctas para mantener el descampado en observación. "Supongamos entonces que ésta es nuestra mejor hipótesis candidata", dije. "¿Podemos pensar en buscar algo que arroje alguna luz especial sobre ella?" Entre los machos, la competencia eclipsa la cooperación más que entre las hembras. ¿Se molestaría un macho en emitir el MIO si su única compañía entre los arbustos fuera otro macho? Seyfarth tuvo una idea mejor: supongamos que un macho produjera el gruñido MIO; ¿sería un macho rival lo bastante tortuoso como para dar una respuesta MIO peligrosamente equívoca cuando viera que el Emisor Original estaba por meterse en un lío? La probabilidad de conseguir alguna vez una buena prueba de esto es minúscula, puesto que habría que observar un caso en el cual el Emisor Original no viera y el Contestador sí viera un depredador próximo *y* el Contestador viera que el Emisor Original no había visto al depredador. (De otro modo el Contestador podría malgastar su credibilidad y provocar la ira y la desconfianza del Emisor Original sin ningún beneficio.) Una coincidencia de condiciones así debe ser sumamente rara. Esta parecía ser una oportunidad ideal para una treta de Sherlock Holmes.

Seyfarth sugirió que podríamos tal vez armar una trampa con algo parecido a una pitón de felpa que le revelaríamos, muy astuta y subrepticiamente, a uno solo de los dos machos que parecían estar por aventurarse fuera de un arbusto. Evidentemene los problemas técnicos serían desagradables y, en el mejor de los casos, sería una conjetura arriesgada, pero con

suerte podríamos conseguir atraer a un embustero hacia nuestra trampa. Pero, al pensarlo mejor, los problemas técnicos parecieron virtualmente insuperables. ¿Cómo podríamos afirmar que el "embustero" había visto realmente (y había sido atrapado por) el depredador y no estaba informando de manera inocente y sincera que la cosa estaba libre? Me sentí tentado (como a menudo antes en nuestras discusiones) a permitirme una fantasía: si yo fuera lo bastante pequeño como para disfrazarme con un traje de mico o si sólo pudiéramos introducir un mico amaestrado o un mico robot o títere que pudiera..." y lentamente me di cuenta de que esta huida recurrente de la realidad tenía un sentido. En realidad no hay ningún sustituto, en la traducción radical de entrar y *conversar con los nativos*. Se pueden verificar más hipótesis en media hora de cháchara tentativa que lo que se puede hacer en un mes de observación y manipulación no entrometida. Pero para sacar provecho de esto hay que volverse entrometido: usted —o su títere— tienen que tener encuentros comunicativos con los nativos, aunque más no sea para andar señalando cosas y preguntando "¿Gavaga?" en un intento por entender lo que significa "Gavagai". Del mismo modo en su travesura de cuento de misterio, alguna parte crucial de la instalación de la trampa del "método Sherlock Holmes" es —*debe ser* lograda por medio de alguna información (equivocada) dada verbalmente. Maniobrar con sus sujetos hasta llevarlos al estado mental adecuado —y saber que se ha tenido éxito— sin la lujosa eficiencia de las palabras puede demostrar ser muy arduo en el mejor de los casos y con frecuencia casi imposible.

En especial, es a menudo casi imposible establecer en el terreno que determinados monos han estado resguardados de cierta información especial. Y puesto que muchas de las hipótesis teóricamente más interesantes dependen exactamente de esas circunstancias, es a menudo muy tentador pensar en trasladar los monos a un laboratorio, donde se pueda *apartar* físicamente a un mono del grupo y darle oportunidades de adquirir información que los otros no tienen *y que el mono de prueba sabe que no tienen*. Esos son los experimentos que Seyfarth y Cheney están llevando a cabo con un grupo de micos cautivos en California y otros investigadores con chimpancés. Los primeros resultados son muy interesantes si bien equívocos (por supuesto), y *tal vez* el ambiente del laboratorio con sus cabinas de aislamiento será la herramienta exacta que necesitamos para abrir la mente de los monos, pero mi presentimiento es que estar aislados de esa manera es un predicamento tan desusado para los micos que probarán no estar preparados por la evolución para sacar provecho de él.

Lo más importante que creo haber aprendido de observar realmente a los micos, es que viven en un mundo en que los secretos son virtualmente imposibles. Al contrario de los orangutanes que son muy solitarios y se juntan sólo para acoplarse y cuando las madres están cuidando la cría, y al contrario de los chimpancés que tienen una organización social fluida en la que los individuos van y vienen viéndose los unos a los otros con bastante frecuencia pero también aventurándose solos mucha parte del tiempo, los micos viven a campo abierto en estrecha proximidad con los otros miembros de sus grupos y no tienen ningún proyecto solitario de ningún alcance. De manera que es en realidad una oportunidad rara cuando un mico está en condiciones de

aprender algo que solamente él sabe *y que sabe que sólo él lo sabe*. (El conocimiento de la ignorancia de los otros y la posibilidad de mantenerla es crítica. Aun cuando un mono es el primero en divisar a un depredador o a un grupo rival, y lo sabe, casi nunca está en posición de estar seguro de que los otros no harán pronto el mismo descubrimiento.) Pero sin que abunden esas ocasiones, hay poco que impartirles a los otros. Más aun, sin tener oportunidades frecuentes de *reconocer* que uno sabe algo que los demás no saben, las razones tortuosas en favor o en contra de impartir información, ni siquiera pueden existir, y menos aun ser reconocidas y actuadas. No puedo pensar en ninguna manera de describir esta simplicidad crítica en el *Umwelt* de los micos, este ingrediente ausente que no se vale explícita o implícitamente de los modismos intencionales de orden más alto.

En suma, los micos no podrían en verdad hacer uso de la mayoría de los rasgos de un idioma humano puesto que su mundo —hasta se podría muy bien decir su estilo de vida— es demasiado simple. Tienen pocas pero intensas necesidades comunicativas, y sus oportunidades para la comunicación son limitadas. Como recién casados que no han estado uno fuera de la vista del otro varios días, se encuentran con que no tienen mucho que decirse (o que decidan guardarse). Pero si no pudieran hacer uso de un lenguaje fantástico, semejante al humano, podemos estar completamente seguros de que la evolución no los ha provisto de éste. Por supuesto que *sí* la evolución los proveyó de un lenguaje complejo con el cual comunicarse, el lenguaje mismo cambiaría radicalmente su mundo y les permitiría crear y transmitir secretos, tan profusamente como lo hacemos nosotros. Y entonces podrían pasar a usar su lenguaje, como nosotros usamos el nuestro en cientos de maneras diversas y marginalmente "útiles". Pero sin los potenciales de información necesarios para cargar la bomba evolutiva, no se podría fijar un lenguaje así.

De manera que podemos estar seguros de que el gruñido MIO, por ejemplo, no está traducido en forma tersa y apropiada por *ningún* intercambio humano conocido. No puede ser una orden (pura, perfecta), ni un pedido, ni una pregunta, ni una exclamación porque no forma parte de un sistema lo bastante complejo como para hacer lugar para esas diferencias sutiles. cuando usted le dice: "¿quieres salir a dar una vuelta?" a su perro y éste salta y emite un ladrido vivaz y menea la cola ante la expectativa, no hay en realidad una pregunta y una respuesta. El perro dispone de solamente unas pocas maneras de "responder". No puede hacer nada equivalente a decir: "Preferiría esperar hasta el atardecer" o "No, si vas a cruzar la autopista" y hasta "No, gracias". Su elocución es una pregunta *en inglés* pero una especie de mezcla derretida de pregunta, orden, exclamación y mero *presagio* (usted ya ha producido esos ruidos relacionados con la salida en otra oportunidad) para su perro (Bennett, 1976, 1983). El gruñido MIO de los micos es, sin duda, una mezcla parecida, pero si bien eso significa que no tendríamos que tener grandes esperanzas de aprender el lenguaje de los micos y averiguar todo acerca de la vida de los monos mediante conversaciones con los micos, no excluye para nada la utilidad de estas hipótesis de traducción algo fantasiosas como maneras de interpretar —y descubrir— los verdaderos papeles o funciones informativas de esas vocalizaciones. Cuando se piensa en el MIO

como "¿Hay moros en la costa?" nuestra atención se dirige hacia una variedad de hipótesis verificable acerca de mayores relaciones y dependencias que deberían ser descubribles si es lo que el MIO significa, o siquiera significa "más o menos".

Reconsideración del paradigma panglossiano

En el contexto del *BBS*, mi codo en defensa del uso de la presunciones de optimalidad por los adaptacionistas y discutiendo la relación de esa táctica con la actitud intencional parecía ser una digresión, que planteaba temas laterales que habría sido mejor haber dejado para otra ocasión. Como lo aclararán estas reflexiones y el próximo capítulo, sin embargo, aquella introdujo un tema central de mi análisis de la actitud intencional: los problemas de interpretación en psicología y los problemas de interpretación en biología son *los mismos problemas* que engendran las mismas perspectivas —y falsas esperanzas— de solución, las mismas confusiones, las mismas críticas y discusiones. Es una afirmación voluminosa, verdadera o falsa, y hay una tendencia a subestimar sus ramificaciones. Algunos la consideran como una verdad relativamente superficial y obvia y adhieren a ella sin darse cuenta de que entonces tienen que renunciar a ser congruentes; otros la descartan casi tan prematuramente, sin admitir que las premisas "obvias" de las que fluyen sus críticas están ellas mismas puestas en tela de juicio por la afirmación. Los malos entendidos más reveladores estuvieron bien representados en los comentarios de *BBS* y corregidos en mi réplica, de la que se extraen los siguientes comentarios, con agregados y correcciones.

El paralelo más importante que quise trazar es éste: los psicólogos no pueden realizar su trabajo sin la asunción de racionalidad de la actitud intencional y los biólogos no pueden hacer el suyo sin las asunciones de *optimalidad* del pensamiento adaptacionista, aunque algunos representantes de cada campo se sientan tentados a negar y censurar el uso de estas presunciones. Las presunciones de optimalidad son tácticas populares en muchas disciplinas, y es a duras penas polémico afirmar que para bien o para mal las ciencias sociales están invadidas por las adopciones de la actitud intencional, algunas mucho más acosadas por problemas que otras. Los debates acerca de las restricciones de *Verstehen* en *Geisteswissenschaft*, la falsa conciencia e ideología en antropología y teoría política, el individualismo metodológico y la clase de funcionalismo de los antropólogos, el papel adecuado de la idealización en economía, el "principio de caridad" en la interpretación y la traducción, todas estas controversias y más son problemas acerca de la justificación para las adopciones especiales de la actitud intencional, y el extraño papel de las formas de la presunción de optimalidad o racionalidad en todos ellos.

En biología, los adaptacionistas dan por sentada la optimalidad del diseño en los organismos que estudian, y esta práctica es mirada con desdén por algunos otros biólogos, puesto que les parece invocar un optimismo doctrinario. ¿Por qué tendría alguien que suponer, en la actualidad, que un organismo, sólo porque ha evolucionado, está diseñado en forma óptima en

algún aspecto? Existe ahora una montaña de pruebas y de buena teoría en la genética de los organismos, por ejemplo, como para demostrar que en muchas condiciones las inadaptaciones están fijadas y las restricciones en el desarrollo limitan la plasticidad genotípica. Pero este desafío está mal planteado; los críticos que les recuerdan a los adaptacionistas estas complicaciones ya están hablando sin oposición.

Esto aparece muy claramente en el comentario de Ghiselin (1983). "La alternativa" para el panglossianismo, dice, "es rechazar por completo esa teleología. En lugar de preguntar: "¿Qué es bueno?", preguntamos: "¿Qué pasó?". La nueva pregunta hace todo lo que podíamos esperar que la vieja hiciera y mucho más" (pág. 362). Esta es una alusión exactamente paralela a la conocida afirmación de Skinner de que la pregunta, "¿cuál es la historia del refuerzo?" es un gran progreso con respecto a "¿Qué cree, quiere, se propone esta persona?". No podemos esperar contestar ninguna de las dos clases de pregunta histórica con una investigación "pura" (completamente no-interpretacional) es decir, sin una porción saludable de presunciones adaptacionistas (o intencionalistas). Esto se debe a que sin respuestas para las preguntas "por qué", no podemos empezar a categorizar *lo que pasó* en partes correctas. El biólogo que se sirve hasta una categoría tan evidentemente segura como *ojo, pierna o pulmón* ya está comprometido con esas asunciones acerca de lo que es bueno, del mismo modo que el psicólogo que se sirve las categorías blandas de *evasión* o *reconocimiento* está comprometido con las presunciones acerca de lo que es racional. (Esto se discute en detalle en el próximo capítulo.)

Adoptamos las presunciones de optimalidad no porque creamos de manera ingenua que la evolución ha hecho de éste el mejor de todos los mundos posibles, sino porque, si hemos de progresar algo, debemos ser intérpretes, y la interpretación requiere la invocación de la optimalidad. [18] Como lo dice Maynard Smith en su comentario (1983), "al hacer uso de la optimización, no estamos tratando de confirmar (o refutar) la hipótesis de que los animales siempre optimizan; estamos tratando de *entender* (con mayor énfasis) las fuerzas selectivas que le dieron forma a su conducta".

La estrategia del adaptacionista en biología busca contestar las preguntas "por qué" exactamente de la misma manera en que lo hace la estrategia intencional en psicología. ¿Por qué, pregunta el psicólogo popular, rechazó John la invitación a la fiesta? La presunción es que hay una (buena)

[18] Kitcher (1987) señala que "es un logro mayor dividir la corriente de la conducta animal en unidades significativas, para describir lo que el animal hizo", pero pasa después a preguntar: "¿Hasta qué punto las descripciones sustentadas por los análisis de optimidad tienen mayores probabilidades de ser ciertas, *sólo en virtud de ese hecho?*". Esta ya es, sutilmente, la pregunta incorrecta a formular. Su doble en psicología sería: ¿Hastà qué punto está apoyada una atribución intencional por una justificación racional con mayores probabilidades de ser cierta, sólo en virtud de ese hecho?; como si existiera la perspectiva de mantener separadas las cuestiones de la atribución y de la justificación racional. Kitcher pregunta cuándo un argumento de optimidad podría lograr "incrementar la probabilidad previa de algunas descripciones funcionales", pero —como para continuar el paralelo en una perspectiva ligeramente diferente— Quine (1960) insistiría en que sería un error preguntar en qué condiciones un hallazgo de consistencia lógica podría "incrementar la probabilidad previa" de una traducción radical hipotética.

razón, por lo menos para John. ¿Por qué, pregunta el adaptacionista, estos pájaros ponen cuatro huevos? El adaptacionista empieza con la suposición de que hay una (buena) razón: que cuatro huevos son mejores, en cierto modo, que dos o tres o cinco o seis. Buscar respuestas posibles para ese "por qué" abre una indagación. En efecto, uno pregunta: ¿si cinco eran demasiados y tres eran demaiado pocos, cuál tendría que ser el caso? Los cálculos hipotéticos se sugieren —acerca del gasto de energía, probabilidad de supervivencia, escasez de comida y demás— y bien pronto, en la buena manera panglossiana, se tiene una explicación candidata de por qué resulta ser para bien que estos pájaros pongan exactamente cuatro huevos, si en realidad eso es exactamente lo que hacen. Por supuesto que puede resultar que, debido a restricciones del desarrollo, las alternativas para los cuatro huevos —como la alternativa de las cuatro patas para los caballos— sean prohibitivamente costosas y, por lo tanto, opciones virtualmente "impensables", pero hasta este descubrimiento se iluminaría si los adaptacionistas plantearan el tema.

Como insisten Gould y Lewontin, siempre se puede tramar una historia así, por tanto la creación de una historia verosímil no es ninguna prueba de que sea verdad. Pero al igual que las mentiras, esas historias ramifican y, o conducen a predicciones falsificadas en otros puntos o no. Si se ramifican sin obstinación, en realidad esto les dice muy poco a los biólogos. "¡Qué maravillosa es la naturaleza!", pueden murmurar desconsoladamente, pero no habrán aprendido mucho. Si, por el contrario, alguna de sus predicciones demuestra ser falsa, los adaptacionistas pueden formular la hipótesis de que se ha omitido algo importante en la explicación. ¿Qué factor de perturbación se podría agregar a la historia panglossiana de manera que lo que los organismos hacen *de verdad*, es, después de todo, lo más sabio para ellos Se empieza con una comprensión ingenua del "problema" que algún organismo enfrenta, y en los términos de esa comprensión ingenua, se resuelve cómo habría que diseñar el organismo. Esto sugiere experimentos que demuestren que el organismo no está diseñado así. En lugar de encogerse de hombros y deducir un diseño "de segunda clase", el adaptacionista pregunta si los resultados señalan una comprensión más elaborada.

Partiendo de la actitud intencional, los psicólogos pueden hacer lo mismo. Se puede encontrar un buen ejemplo en la investigación reciente de Kahneman y Tversky (1983). Le formularon una pregunta a su grupo de estudios. Supongan que ustedes habían comprado una entrada para el teatro por diez dólares y que cuando llegaron al teatro se dieron cuenta de que habían perdido la entrada. Hay más entradas en venta y ustedes tienen suficiente dinero en la billetera. ¿Gastarían diez dólares para reemplazar la entrada que perdieron? Más de la mitad de los miembros del grupo expresan la convicción de que no comprarían otra entrada. A otros se les pide que consideren esta variante. Usted planea ir al teatro y cuando va a la taquilla a comprar la entrada, descubre que ha perdido un billete de diez dólares. ¿Compraría la entrada lo mismo? Sólo el doce por ciento de los sujetos dicen que la pérdida de los diez dólares influiría sobre su decisión de adquirir una entrada.

Hay una diferencia clara en las respuestas a las dos preguntas, pero Kahneman y Tversky señalan que las dos circunstancias son equivalentes. En

ambos casos, si usted va al teatro tiene diez dólares menos al final del día que si no va. ¡Seguramente es irracional tratar los dos casos de distinta manera! No hay duda de que un prejuicio así es racionalmente indefendible si todo lo que tomamos en cuenta son los costes y los beneficios mencionados hasta ahora. ¿Debemos inferir entonces que esto no es *nada más* que un indicio de la deplorable fragilidad humana? ¿O hay alguna otra perspectiva desde la cual defender este prejuicio? Kahneman y Tversky sugieren que si dividimos los "gastos de contabilidad" podemos descubrir beneficios en la eficiencia y la lucidez del planeamiento si dividimos nuestra riqueza en "cuentas" separadas y distinguimos claramente entre categorías tales como las "pérdidas" y "el coste de hacer negocios" (por ejemplo). Estos beneficios pueden valer mucho más que las pérdidas ocasionales que sufrimos cuando esa política contable dicta una decisión por debajo de lo óptimo. La hipótesis de que esos gastos de contabilidad podrían importar tanto sugiere algo acerca de la organización interna de nuestros sistemas cognitivos. Nada está probado (todavía), pero el descubrimiento de esta debilidad *sistemática* en nuestro modo normal de razonar acerca de esos temas sugiere que su eliminación sería más costosa que lo que podríamos habernos imaginado.

Lo anterior ejemplifica el paralelo que yo quería trazar entre la teorización adaptacionista e intencionalista, pero hay otra manera de alinear las disciplinas que les ha sugerido a varios autores que tengo la analogía exactamente al revés. Dahlbom (1985) adopta el punto de vista del pájaro.

> En estos días el romanticismo es moda en la ciencia. Las ideas profundamente arraigadas que son el centro de nuestra tradición iluminista están siendo cuestionadas. Hay una tendencia alejada del atomismo, del empirismo, del funcionalismo (adaptacionismo) y del gradualismo y hacia el holismo, el innatismo, el estructuralismo y el transicionismo... El estudio crítico de Chosmky del *Verbal Behavior* de Skinner (Chomsky, 1959) fue una primera publicación gráfica admirable que anunciaba la nueva tendencia.
>
> La conmoción reciente en la teoría evolutiva provocada por Eldredge, Gould, Lewontin, Stanley y otros... es sólo otro ejemplo de esta tendencia romántica. ¿Entonces por qué Dennett elige colocar a Gould y Lewontin en el mismo campo con Skinner, justamente con él, más que con Chomsky, Kuhn y otros, donde pertenecen claramente? (pág. 760).

Amundson (inédito) describe la misma línea de batalla en función de teorías que "explican los caracteres existentes como resultados ambientales seleccionados de una variación generada al azar" *versus* las teorías según las cuales el "ambiente puede *dar forma* al resultado expreso del desarrollo pero los efectos formadores están significativamente constreñidos por la estructura interna".

> Las teorías del primer grupo (llamémoslas "ambientalistas") incluyen teorías de aprendizaje conductista y la biología seleccionista adaptacionista. Las teorías del segundo tipo ("estructuralistas") incluyen la psicología innatista y cognitiva y las teorías evolucionistas que recalcan, por ejemplo, las represiones morfológicas y embriológicas de la evolución.

La figura pivote de esta analogía, como lo sugiere Dahlbom, es Chomsky cuyo innatismo extremo equivalente a la negación de que existe algo tal como el aprendizaje representa el lado oscuro —romántico— de la ciencia cognitiva. (Una vez le sugerí a Chomsky que de acuerdo con su punto de vista nadie *aprendía* nunca mecánica cuántica pero que algunas personas, gracias a la estructura de su talento innato "grababan" la mecánica cuántica en la memoria. El estuvo de acuerdo.) Y, bastante curiosamente este estructuralismo extremo tiene, en verdad, una cosa en común con el conductismo extremo: una voluntad prematura de dejar de preguntar "por qué". ¿Por qué todas las gramáticas tienen una característica tal y tal? Porque ésa es la manera en que están construidos los usuarios del lenguaje. Fin de la explicación. ¿Por qué están construidos de ese modo? ¿Por qué surgieron esas estructuras restrictivas? En opinión de Chomsky ésa no es una pregunta para ser contestada por lingüistas o psicólogos.

No obstante, hay que darle alguna explicación a toda esa estructura, y como yo argumenté en "Passing the Buck to Biology" (1980a), es una estrategia completamente razonable para examinar esos procesos próximos y accesibles antes de optar por alternativas más distantes. "Cuanto más se pueda considerar al cerebro del bebé como una *tábula rasa*, más accesibles a la investigación *experimental* serán los *últimos* misterios del aprendizaje. Si los hechos constriñen a los psicólogos a pasarles la responsabilidad a los biólogos evolucionistas, tendremos que transar con respuestas más abstractas y especulativas a las preguntas finales." (Pág. 19).

¿Sin embargo, a quién pueden cargarle la responsabilidad los teóricos estructuralistas de la biología evolutiva? ¿Qué explica la existencia del *Pauplane* represor en relación con el cual las adaptaciones son sólo una sintonización refinada? Se puede decir: la biosfera está construida exactamente de esa manera. ¿Fin de la explicación? Tal vez, pero aquellos de nosotros que buscamos el esclarecimiento siempre estaremos listos para preguntar, una vez más, ¿por qué? Es el desagrado puritano por este pensamiento teleológico y funcional lo que une —en este sentido— no sólo a Skinner y Ghiselin sino también a Lewontin y Chomsky, dejando a Kahneman, Tversky, Dawkins, Maynard Smith y otros intencionalistas del otro lado del cerco.

El ataque de Gould y Lewontin al pensamiento adaptacionista, y la polémica que provocó han sido instructivos a pesar de —y por cierto a veces debido a— la manera en que los participantes se han sentido tentados a hablar sin importarles los demás. Como Kitcher (1985) lo señala en su resumen del episodio:

> Gould y Lewontin hacen campaña para atraer la atención hacia formas rivales de hipótesis evolucionistas, pero distorsionan su punto de vista al sugerir que la imposibilidad de falsificar las afirmaciones adaptacionistas es un obstáculo insuperable. Si yo estoy en lo cierto, la posición correcta es que el seguimiento con éxito de las hipótesis adaptacionistas acerca de las características de los organismos ya presupone exactamente esa atención a las posibilidades rivales que Gould y Lewontin recomiendan a sus colegas (pág. 232).

Mayr (1983) da un veredicto coincidente.

> Parece evidente que el programa adaptacionista como tal tiene poco que sea incorrecto, contrariamente a lo que alegan Gould y Lewontin, pero que no debería ser aplicado de una manera atomista exclusiva. No hay mejor prueba para esta conclusión que aquella que Gould y Lewontin mismos han presentado. Las preguntas "por qué" aristotélicas son muy legítimas en el estudio de las adaptaciones, siempre que se tenga una concepción realista de la selección natural y se comprenda que el individuo como un todo es un sistema genético y de desarrollo complejo y que llevará a respuestas absurdas si se hace añicos este sistema y se analizan uno por uno los despojos (pág. 332).

Estimulado, entonces, por lo que yo considero una muy buena compañía (véase también Rosenberg, 1985) me planto en mis opiniones:

1) El pensamiento adaptacionista en biología es precisamente tan inevitable, tan provechoso —y tan riesgoso— como el pensamiento mentalista en psicología y la ciencia cognitiva en general.

2) El pensamiento adaptacionista propiamente dicho *está* adoptando tan sólo ahora una versión especial de la actitud intencional en el pensamiento evolutivo, descubriendo las "razones de ser de flotación libre" de los diseños en la naturaleza. [19]

El adaptacionismo como interpretación radical retrospectiva

En realidad este punto de vista no es tan radical como les parece a algunos. Puede ser sorprendente, pero ha estado implícito todo el tiempo en los triunfos incontestables de la biología darwiniana (y neodarwiniana), tal como Dawkins lo muestra en detalle en *The Blind Watchmaker* (1986). A pesar de la ortodoxia de esta posición, sin embargo, sigue inquietando a algunos. Las fuentes de la resistencia a este punto de vista incluyen una variedad sorprendente de ideologías y fobias, además de aquellas analizadas en este capítulo; y en el capítulo 8 demostraré como una curiosa constelación de filósofos —Searle, Foder, Dretske, Burge y Kripke— están unidos por su antipatía por algunas de sus implicaciones. En preparación para eso, quiero señalar un problema especial de la prueba que la teoría evolutiva enfrentó. Como lo han señalado muchos comentaristas, las explicaciones evolucionistas son esencialmente narraciones históricas. Mayr (1983) lo expresa así: "Cuando uno intenta explicar las características de algo que es un producto de la evolución, tiene que intentar reconstruir la historia evolutiva de esta característica" (pág. 325). Pero, como veremos, determinados hechos históricos juegan un papel evasivo en esas explicaciones.

[19] Discusiones aclaratorias de otras facetas de la relación entre la teoría evolucionista y la psicología —discusiones con las que no siempre estoy de acuerdo— se encuentran en Patricia Kitcher (1984), Sober (1985) y Rosenberg (1986a, b).

La teoría de la selección natural muestra cómo cada característica del mundo natural *puede* ser el producto de un proceso finalmente mecánico, ciego, imprevisible, noteológico de la reproducción diferencial durante largos períodos. Pero, por supuesto, a algunos rasgos del mundo natural —las patas cortas de los perros salchicha y del ganado vacuno Black Angus, la piel gruesa de los tomates— son el producto de la selección artificial en la cual el objetivo del proceso y la razón de ser de los diseños a los que se aspira, jugaron un papel explícito en la etiología, al estar "representados" en las mentes de los criadores que hicieron la selección. Por tanto, entonces, la teoría de la selección natural debe hacer lugar para la existencia de esos productos y esos procesos históricos como casos especiales. ¿Pero se pueden distinguir esos casos especiales en el análisis retrospectivo? Téngase en cuenta un experimento del pensamiento extraído en mi réplica al comentario en *BBS*.

Imaginemos un mundo en el cual manos *reales* suplementaran a la "mano oculta" de la selección natural, un mundo en el que la selección natural hubiera sido ayudada y encubierta por sobre los eones por diseñadores de organismos chapuceros, previsores, representantes de la razón, como los criadores de animales y plantas de nuestro mundo real, pero que no se limitaban a los organismos "domesticados", diseñados para el uso humano. Estos bioingenieros habrían formulado y representado y actuado de verdad sobre la razón de ser de sus diseños, exactamente como los ingeieros del automóvil. Ahora bien, ¿sería detectable la obra de sus manos por los biólogos de ese mundo?. ¿Sus productos se podrían distinguir de los productos de un zarandeo "puramente" darwiniano, sin agente, sin representatividad, donde todas las razones de ser eran de flotación libre? Por supuesto que sí (por ejemplo, si algunos organismos vinieran con manuales de mantenimiento incluidos), pero no se podrian distinguir, si los ingenieros decidieran ocultar su intervención lo mejor que pudieran. [20]

¿Revelaría una mirada más atenta a los diseños algunas discontinuidades delatoras? La selección natural, al carecer de previsión, no ve la prudencia de *reculer pour mieux sauter*: dar un paso atrás para saltar mejor hacia adelante. Si hay diseños a los que no es posible acercarse por medio de un proceso de rediseño gradual, escalonado, en el cual cada paso por lo menos no es peor para las probabilidades de supervivencia de los genes que su predecesor, entonces, la existencia de un diseño así en la naturaleza, parecería requerir, en algún punto de su evolución, la ayuda de un diseñador previsor

[20] Nova-Gene, una compañía biotecnológica de Houston, ha adoptado la política de "el marcaje ADN: escribiendo la interpretación más aproximada de la marca registrada para los aminoácidos de su compañía en el ADN "extra" o "de desecho" de sus productos (según las abreviaturas estándar, ácido asparagina-glutamina-valina-alanina-glicina-glutámico-ácido asparagina-glutámico = NQVAGENE) (Scientific American, julio 1986, págs. 709-71). Esto sugiere un nuevo ejercicio en la traducción radical para los filósofos: ¿cómo podríamos confirmar o no la hipótesis de que las marcas registradas —o los manuales de mantenimiento— serían perceptibles en la masa del ADN que aparentemente no está involucrado en la dirección de la formación del fenotipo? El punto de vista del ojo del gen de Dawkins predice, y por tanto podría explicar la presencia de este "ADN egoísta", sin sentido (véase Dawkins, 1982, capítulo 9: "El ADN egoísta, los genes saltones y un susto lamarckiano") pero eso no demuestra que *no podría* tener un origen más dramático, y por tanto un significado, después de todo.

—ya sea un empalmador de genes o un criador que de alguna manera preservara a la necesaria sucesión de reincidentes intermedios hasta que pudieran rendir su buscada progenie—. ¿Pero semejante salto hacia adelante —una "transición" en el lenguaje de los teóricos evolucionistas— ¿no podría ser un mero salto afortunado? ¿En qué punto rechazamos las hipótesis del accidente cósmico como demasiado improbable y aceptamos las hipótesis de los ingenieros intervencionistas? [Véanse las discusiones de gradualismo, transición y probabilidad en Dawkins (1986).]

Estas preguntas sugieren —pero por supuesto no prueban— que podrían no haber señales inequívocas de selección natural (como opuesta a la artificial). Si se probara esta conclusión ¿sería ella una vergüenza para los evolucionistas en su lucha contra los creacionistas? Es posible imaginar el griterío: Los científicos lo admiten: ¡"La teoría de Darwin no refuta al Designio Inteligente!". Pero esto es confundir el status de la teoría evolucionista ortodoxa. Sería sumamente arriesgado para cualquier defensor de la teoría de la selección natural proclamar que ella nos otorga el poder de estudiar la historia tan sutilmente a partir de los datos actuales como para excluir del todo la presencia histórica anterior de los diseñadores racionalistas. Sería una fantasía por completo sin pies ni cabeza, inverosímil, pero, después de todo, una posibilidad.

En nuestro mundo actual hay organismos que *sabemos* que son el resultado de los esfuerzos de re-diseño previsores, perseguidores de un objetivo, pero esa certeza depende de nuestro conocimiento directo de hechos históricos recientes (hemos estado observando de verdad a los criadores en su trabajo), y estos hechos especiales podrían no proyectar ninguna sombra fósil sobre el futuro. Como para tomar en cuenta una variación más sencilla en nuestro experimento del pensamiento, supongamos que les mandáramos a los biólogos marcianos una gallina ponedora, un perro pekinés, una golondrina y una chita y les pidiéramos que determinaran qué diseños llevaban la marca de la intervención de seleccionadores artificiales. ¿En qué podrían confiar? ¿Cómo argumentarían? Podrían darse cuenta de que la gallina no cuidaba "bien" a sus huevos; algunas variedades de gallina han sido despojadas de su instinto maternal y estarían en vías de extinción si no fuera por el ambiente de las incubadoras artificiales que los seres humanos les han brindado. Podrían notar que el pekinés estaba patéticamente mal preparado para valerse por sí mismo en cualquier ambiente exigente. La inclinación de la golondrina por los nidos de carpintería podría hacerles pensar que era una especie de mascota y cualquier característica del chita que los hubiera convencido de que era un ser salvaje podría encontrarse también en los perros galgos y haber sido estimulada pacientemente por los criadores. Después de todo, los ambientes artificiales son una parte de la naturaleza.

El toqueteo prehistórico de parte de los visitantes intergalácticos con el ADN de las especies terráqueas no se puede descartar, excepto sobre la base de que es una fantasía enteramente gratuita. Nada de lo que hemos encontrado (hasta ahora) en la Tierra insinúa siquiera que esa hipótesis sea digna de una investigación mayor. (Y, obsérvese —me apresuro a agregar, para que los creacionistas no se envalentonen— que aun si fuerámos a descubrir y traducir un "mensaje de marca de fábrica" a nuestro ADN de reserva, esto

no serviría para anular la afirmación de la teoría de la selección natural para explicar todos los diseños de la naturaleza sin invocar a un Creador-Diseñador precavido de *fuera del sistema*. Si la teoría de la evolución por selección natural puede explicar la existencia de la gente de Nova Gene que inventó el marcaje ADN, también puede explicar la existencia de antepasados que pueden haber dejado sus firmas cerca para que nosotros las descubriéramos.) El poder de la teoría de la selección natural no es el poder para probar con exactitud cómo era la (pre) historia sino sólo el poder para probar cómo podría haber sido, dado lo que sabemos acerca de cómo son las cosas.

El pensamiento adaptacionista, entonces, puede a menudo ser incapaz de contestar preguntas especiales acerca de rasgos específicos de los mecanismos históricos, la etiología real de una evolución del diseño natural, si bien puede lograr formular y hasta confirmar —hasta donde la confirmación fuera posible— un análisis funcional del diseño. La diferencia entre que un diseño tenga una razón de ser de flotación libre (no representada) en su historia y que tenga una razón de ser representada, puede muy bien ser imperceptible en las características del diseño, pero esta incertidumbre es independiente de la confirmación de esa razón de ser para ese diseño. Más aun, como veremos en el próximo capítulo, los hechos históricos acerca del proceso de evolución del diseño, aun cuando podamos descubrirlos, son igualmente neutrales cuando nos movemos en la dirección contraria: son incapaces de resolver las preguntas acerca de la razón de ser del diseño de las que depende nuestra interpretación de sus actividades. Podríamos todavía confiar en que la ciencia eventualmente descubrirá la verdad histórica acerca de estos detalles etiológicos, pero no porque vaya a resolver todas nuestras preguntas "por qué" aristotélicos, aun cuando se las plantee con cautela y propiedad.

8

La evolución, el error y la intencionalidad*

A veces les lleva años de debate a los filósofos descubrir acerca de qué están en desacuerdo, en realidad. A veces hablan con total olvido de los demás en largas series de libros o artículos, sin adivinar nunca la raíz del desacuerdo que los divide. Pero oportunamente llega el día, algo parece hacer que el gato salga de la bolsa. "¡Ajá!", exclama un filósofo ante otro, "¡así que ésa es la razón por la cual has estado disintiendo conmigo, entendiéndome mal, resistiendo mis conclusiones, desconcertándome todos estos años!".

En el otoño de 1985 descubrí lo que tomé por ser nada más que un desacuerdo sumergido —tal vez hasta reprimido— y supuse que harían falta algunas tácticas de choque para impulsar ese secreto embarazoso hasta la severa e iracunda mirada de la atención filosófica. Hay pocas cosas más chocantes para los filósofos que la alianza de personas incompatibles, por tanto, en un borrador anterior de este capítulo que circuló ampliamente en 1986, tracé algunas líneas de batalla deliberadamente supersimplificadas y tomé partido: los buenos contra los malos. Resultó. Aquellos a quienes yo había desafiado y otros que mordieron el anzuelo me inundaron con respuestas detalladas sumamente reveladoras. De una manera general, estas reacciones confirmaron tanto mi división del campo como mis afirmaciones en favor de su no reconocida importancia.

Sin embargo, las respuestas fueron tan constructivas aun de parte de aquellos a quienes había tratado en forma más bien ruda —o representando mal— en el primer borrador que en lugar de cacarear "¡se lo había dicho!", debería reconocer desde el principio que este fruto de mi primer acto de provocación intensamente corregido y expandido tiene una deuda especial con los comentarios de Tyler Burge, Fred Dretske, Jerry Fodor, John Haugeland, Saul Kripke, Ruth Millikan, Hilary Putnam, Richard Rorty, y Stephen Stich, y con muchos otros, incluyendo especialmente a Fred Adams, Peter Brown, Jerome Feldman, D. K. Modrak, Carolyn Ristau, Jonathan Schull, Stephen White, y Andrew Woodfield.

La gran división que quiero exponer resiste una formulación simple y directa, lo que no es sorprendente, pero podemos situarla desandando los pasos de mi indagación que comenzó con un descubrimiento acerca de las acti-

* Todo menos la última sección de este capítulo aparece con el mismo título, en Y. Wilks y D. Partridge, eds., *Source Book on the Foundations of Artificial Intelligence* (Cambridge: Cambridge University Press, 1987) y está vuelto a imprimir con autorización.

tudes de ciertos filósofos hacia la interpretación de los artefactos. Se me cayó la venda de los ojos durante una discusión con Jerry Fodor y algunos otros filósofos sobre un borrador de un capítulo de *Psychosemantics* (1987). Las vendas se me caen a menudo de los ojos cuando discuto cosas con Fodor, pero ésta fue la primera vez, hasta donde puedo recordar, en que realmente me descubrí murmurando "¡ajá!" para mí mismo. El capítulo en cuestión, "Meaning and the World Order", se ocupa de las tentativas de Fred Dretske (1981, especialmente capítulo 8; 1985; 1986) para resolver el problema de la representación errónea. Como ayuda para comprender el tema, yo le había propuesto a Fodor y a los demás participantes de la discusión que discutiéramos primero un caso muy simple de representación fraudulenta: un aparato de verificación de la ranura de una máquina expendedora que aceptaba un pedazo de metal. "Ese tipo de caso es irrelevante", replicó Fodor instantáneamente, "puesto que después de todo, John Searle tiene razón acerca de una cosa; tiene razón acerca de esa clase de artefactos. No tienen ninguna intencionalidad intrínseca u original, sólo una intencionalidad derivada."

La doctrina de la intencionalidad original es la afirmación de que si bien algunos de nuestros artefactos pueden haber derivado intencionalmente de nosotros, tenemos una intencionalidad original (o intrínseca), totalmente no derivada. Aristóteles dijo que Dios es el Mudador inmóvil y esta doctrina anuncia que nosotros somos significadores sin significado. Nunca he creído en ella y a menudo he discutido en su contra. Como lo ha señalado Searle, "Dennett... cree que nada tiene *literalmente* ningún estado mental *intrínsecamente intencional*" (1982, pág.57), y en el prolongado debate entre nosotros (Searle 1980b, 1982, 1984, 1985; Dennett 1980b; Hofstadter y Dennett 1981; Dennett 1982c, 1984b, de próxima aparición f), yo había dado por sentado que Fodor estaba de mi parte en este punto especial.

¿Creía Fodor realmente que Searle tiene razón sobre esto? El dijo que sí, Dretske (1985) llega más lejos citando el ataque de Searle a la inteligencia artificial (Searle 1980) con aprobación, y trazando un agudo contraste entre la gente y los ordenadores:

> Carezco de habilidades, conocimiento y comprensión especializadas, pero de nada que sea esencial para formar parte de la sociedad de los agentes racionales. No obstante, con las máquinas, y esto incluye a los ordenadores modernos más complejos, es diferente. Ellas *sí* carecen de algo que es esencial (pág. 23).

Otros que lucharon recientemente con el problema de la representación fraudulenta o el error, también me parecieron ponerse del lado de Searle: en general, Tyler Burge (1986) y Saul Kripke (1982, especialmente pág. 34 y sigs). En realidad, como veremos, el problema del error tortura sólo a quienes creen en la intencionalidad original o intrínseca.

¿Son la *intencionalidad original* y la *intencionalidad intrínseca* la misma cosa?

Tendremos que enfocar esta pregunta en forma indirecta buscando distintas tentativas de trazar una aguda distinción entre la manera en que nuestras mentes (o estados mentales) tienen sentido y la manera en que lo

tienen otras cosas. Podemos comenzar con una distinción conocida e intuitiva discutida por Haugeland. Nuestros artefactos

> ...sólo tienen sentido porque nosotros se los damos; su intencionalidad, como la de las señales de humo y la escritura, son esencialmente prestadas, por tanto *derivativas*. Para expresarlo sin ambages: los ordenadores mismos no significan nada por sus signos (no más que los libros); sólo significan lo que nosotros decimos que significan. Por otra parte, la comprensión auténtica es intencional "por derecho propio" y no derivada de alguna otra cosa (1981, págs.32-33).

Consideremos una enciclopedia.Tiene intencionalidad derivada. Contiene información acerca de miles de cosas en el mundo, pero solamente hasta donde es un artefacto diseñado y destinado para nuestro uso. Supongamos que "automatizamos" nuestra enciclopedia poniendo todos sus datos en una computadora y convirtiendo su índice en la base de un sistema complejo de preguntas y respuestas. Ya no tenemos que buscar material en los volúmenes: simplemente tecleamos preguntas y recibimos respuestas. A los usuarios ingenuos les podría parecer como si se estuvieran comunicando con otra persona, otra entidad dotada de intencionalidad original, pero nosotros no seríamos tan tontos. Un sistema de preguntas y respuestas no es nada más que una herramienta y, cualquiera que sea el significado o acerquidad que le adjudiquemos, no es más que un subproducto de nuestras prácticas al usar el artefacto para servir nuestros propios fines. No tiene objetivos propios, con excepción del objetivo artificial y derivado de "entender" y "contestar" correctamente nuestras preguntas.

Pero supongamos que dotemos a nuestro ordenador de objetivos algo más autónomos, algo menos esclavos. Por ejemplo, un ordenador que juega al ajedrez tiene el objetivo (artificial, derivado) de derrotar a su rival humano, de ocultar lo que "sabe" por nosotros, de trampearnos tal vez. Pero seguramente, sin embargo, es sólo nuestra herramienta o juguete y aunque muchos de sus estados internos tengan una especie de acerquidad o intencionalidad —por ejemplo, hay estados que representan (y por tanto son acerca de) distintas continuaciones posibles del juego— ésta es simplemente una intencionalidad derivada, no original.

Este tema persuasivo (no es realmente un argumento) ha convencido a muchos pensadores de que ningún artefacto podría tener la clase de intencionalidad que nosotros tenemos. Cualquier programa de computación, cualquier robot que pudiéramos diseñar y construir, por más fuerte que fuera la ilusión que podemos crear de que se ha convertido en un agente auténtico, no podría ser nunca un pensador verdaderamente autónomo con la misma clase de intencionalidad original que nosotros disfrutamos. Por el momento, supongamos que ésta es la doctrina de la intencionalidad original y veamos dónde conduce.

El caso del ordenador errante de dos bits

Forzaré ahora mi ejemplo de la máquina expendedora —el ejemplo que Fodor insistió en que no era pertinente— explícitamente porque vuelve vívi-

dos exactamente los puntos de desacuerdo y coloca algunas polémicas recientes (acerca de la "psicología individualista" y el "contenido restringido", acerca del error, acerca de la función) bajo una luz útil. Tómese en cuenta una máquina estándar expendedora de gaseosas diseñada y construida en los Estados Unidos y equipada con un dispositivo transductor para aceptar y rechazar monedas de 25 centavos.[1] Llamemos a ese dispositivo un dispositivo de dos bits. Normalmente, cuando se inserta una moneda de 25 centavos en un dispositivo de dos bits, el dispositivo entra en un estado, llamémoslo *Q*, que "significa" (observe las citas alarmantes) "percibo/acepto una moneda norteamericana auténtica de 25 centavos en este momento". Esos dispositivos son muy inteligentes y elaborados pero en absoluto infalibles. En realidad "cometen errores" (más citas alarmantes). Es decir, que no metafóricamente a veces entran en el estado *Q* cuando se les introduce un trozo de metal u otro objeto extraño, y a veces rechazan monedas perfectamente legales: no entran en el estado *Q* cuando *se supone que deberían* hacerlo. No hay duda de que hay modelos detectables en los casos de "percepción equivocada". No hay duda de que por lo menos algunos de los casos de "falsa identificación" podrían ser pronosticados por alguien con bastante conocimiento de las leyes pertinentes de la física y los parámetros de diseño del mecanismo transductor del dispositivo de dos bits, de manera que fuera igualmente un tema de la ley de la física que los objetos de la clase *K* pusieran al dispositivo en el estado *Q* igual que las monedas de 25 centavos. Los objetos de la clase *K* serían "buenos pedazos de metal", que "engañarían" confiablemente al transductor.

Si los objetos de la clase K se volvieran más comunes en el entorno normal del dispositivo de dos bits, podríamos esperar que los propietarios y diseñadores de esos dispositivos desarrollaran transductores más avanzados y sensibles que discriminaran confiablemente entre las monedas legítimas de 25 centavos y los pedazos de lata de clase K. Por supuesto, podrían aparecer monedas falsas más tramposas que exigieran mayores adelantos en los transductores detectores y en algún punto esa escalada de la ingeniería alcanzaría devoluciones en disminución porque no existe ningún mecanismo *infalible*. Mientras tanto, los ingenieros y usuarios tendrán la prudencia de arreglárselas con dispositivos de dos bits estándar y rudimentarios, puesto que no conviene gastar en protegerse uno mismo de los abusos insignificantes.

La única cosa que hace del dispositivo un detector de monedas de 25 centavos más que de pedazos de lata o un detector de moneda de 25 —o pedazo de lata— es la intención compartida de los diseñadores, constructores, propietarios y usuarios del dispositivo. Es sólo en el ambiente o contexto de esos usuarios y sus intenciones que podemos escoger algunas de las oportunidades del estado *Q* como "verídicas" y otras como "equivocadas". Es sólo en relación con ese contexto de intenciones que podríamos justificar llamar al dispositivo un dispositivo de dos bits en primer término.

Doy por sentado que hasta ahora tengo a Fodor, Searle, Dretske, Burge, Kripke y otros asintiendo. Así es como ocurre con esos artefactos; éste es

[1] Esta táctica es escasamente innovadora. Entre las primeras discusiones de intencionalidad basadas en esos ejemplos de mecanismos de discriminación simple están MacKenzie, inédito (1978), Ackermann, 1972 y Enc 1982.

un caso de libro de texto de intencionalidad derivada totalmente expuesta. Y por tanto, no abochorna a nadie admitir que un dispositivo de dos bits determinado, de procedencia directa de una fábrica norteamericana y con las palabras "dispositivo de dos bits modelo A" grabadas en él podría ser instalado en una máquina de gaseosas panameña donde procedía a ganarse el sustento como aceptador y rechazador de balboas de 25 centavos, moneda legal de Panamá, y fácilmente distinguible de los 25 centavos norteamericanos por el diseño y la inscripción grabadas en ellas pero no por su peso, grosor, diámetro o composición material.

(No estoy inventando esto. Me dio el dato una autoridad excelente —Albert Erler del Flying Eagle Shoppe, monedas raras— de que los balboas panameños de 24 centavos acuñados entre 1966 y 1984 no pueden ser distinguidos de las norteamericanas por las máquinas expendedortas estándar. No es de extrañar, puesto que son producidos sacándolas del stock de monedas de 25 norteamericanas en las casas de moneda de los Estados Unidos. Y —para satisfacer a los curiosos, aunque no es pertinente para el ejemplo— ¡la tasa de cambio oficial corriente para el balboa de 25 centavos es, por cierto, de 25 centavos de dólar!)

Un dispositivo de dos bits así, despachado a Panamá (el Planeta Tierra Gemelo del hombre pobre), todavía entraría en cierto estado físico —el estado que tiene las características físicas por las cuales solíamos identificar al estado *Q*— cada vez que una moneda de 25 centavos norteamericana o un objeto de clase K o un balboa de 25 centavos panameño se inserta en ella, pero ahora un conjunto diferente de tales ocasiones se cuenta como los errores. En el nuevo ambiente las monedas norteamericanas son como pedazos de metal, como inducidoras de error, percepción falsa, representación fraudulenta, tanto como los objetos de la clase K. Después de todo, en los Estados Unidos, un balboa panameño de 25 centavos también es una especie de pedazo de metal.

Una vez que nuestro dispositivo de dos bits se instala en Panamá, ¿deberíamos decir que el estado que solíamos llamar *Q* todavía se produce? El estado físico en el que el dispositivo "acepta" monedas todavía ocurre, ¿pero deberíamos decir ahora que tendríamos que identificarlo como "realizando" un estado nuevo, *QB* en cambio? Pues bien, hay una libertad considerable —por no decir aburrimiento— acerca de lo que tendríamos que decir, puesto que después de todo un dispositivo de dos bits no es más que un artefacto y hablar de sus percepciones y sus falsas percepciones, sus estados verídicos y no verídicos —en pocas palabras su intencionalidad— es "sólo una metáfora". El estado interno del dispositivo de dos bits, llámelo lo que quiera, no significa *en realidad* (originalmente, intrínsecamente) ni "25 centavos estadounidenses aquí y ahora" ni "balboa panameño de 25 centavos aquí y ahora". En *realidad*, no significa nada. Por tanto Fodor, Searle, Dretske, Burge y Kripke (*inter alia*) insistirán.

El dispositivo de dos bits fue diseñado originalmente para ser un detector de monedas norteamericanas de 25 centavos. Esa era su "función adecuada" (Millikan, 1984), y muy literalmente su *raison d'être*. Nadie se hubiera molestado en ponerla en existencia si no se les hubiera ocurrido este propósito. Y

dado que este hecho histórico acerca de su origen permite una cierta manera de hablar, un dispositivo así puede ser caracterizado primaria u originalmente como un dispositivo de dos bits, una cosa cuya función es detectar las monedas de 25 centavos, de manera que *en relación con esa función* podemos identificar tanto sus estados verídicos como sus errores.

Esto no impediría que un dispositivo de dos bits fuera arrancado de su nicho doméstico y forzado a prestar servicio con un nuevo propósito —cualquiera que sea el nuevo propósito que las leyes de la física certifican que podría servir de manera confiable— como un detector de K, un detector de balboas de 25 centavos, un tope de puertas, un arma mortal. En su nuevo papel podría haber un breve período de confusión o indeterminación. ¿Cuán largo debe ser el recorrido que algo tiene que acumular antes de dejar de ser un dispositivo de dos bits para ser más bien un detector de balboas de 25 centavos o un tope de puerta o un arma mortal? En el mismo momento de su debut como detector de balboas de 25 centavos después de diez años de fieles servicios como dispositivos de dos bits, ¿es su estado ya una detección *verídica* del balboa de 25 centavos o podría haber una suerte de error de nostalgia por la fuerza de la costumbre, una identificación equivocada de un balboa de 25 centavos, *como* una moneda norteamericana de 25 centavos?

Tal como se ha descrito, el dispositivo de dos bits difiere notablemente en que no tiene ninguna provisión para el recuerdo de sus experiencias pasadas; ni siquiera una "memoria" (en citas alarmantes) para sus "experiencias" pasadas. Pero por lo menos la última se le podría proporcionar fácilmente si se pensara que tiene importancia. Para empezar con la incursión más simple dentro de este tema, supongamos que el dispositivo de dos bits (para referirnos a él con su nombre de pila original) está equipado con un contador, el que después de diez años de servicio está en 1.435.792. Supongamos que no se lo vuelve al cero durante su vuelo a Panamá, de modo que el día de su debut allí, el contador pasa a 1.435.793. ¿Inclina esto la balanza en favor de la afirmación de que todavía no se ha conectado a la tarea de identificar correctamente los balboas de 25 centavos? ¿Conducirían las variaciones y complicaciones de este tema en direciones diferentes a sus intuiciones?

Podemos estar seguros de que nada *intrínseco* acerca del dispositivo de dos bits considerado apenas por sí mismo o independientemente de su historia anterior lo distinguiría de un auténtico detector de balboas fabricado especialmente por encargo del gobierno panameño. Sin embargo, dada su historia, *¿no hay allí un problema sobre su función, su propósito, su significado*, en esta primera ocasión en que entra en el estado en que estamos tentados de llamar *Q*? ¿Es éste un caso de entrar en el estado *Q* (con el significado de "moneda norteamericana de 25 centavos") o en el estado *QB* (con el significado de "balboa panameño de 25 centavos aquí y ahora")? Yo diría, junto con Millikan (1984) que si su debut panameño cuenta como entrar en el estado *Q* o en el estado *QB* depende de si, en su nuevo nicho, fue *seleccionada por* su capacidad de detectar balboas de 25 centavos, escogida literalmente, por ejemplo, por el licenciatario de la franquicia panameña para Pepsi-Cola. Si fue escogida así, incluso si sus nuevos dueños se hubieran podido olvidar de reajustar su contador, su primer acto "perceptivo" contaría como

una identificación correcta como detector de balboas de 25 centavos, porque es eso *para lo que* se la usaría ahora. (Habría adquirido la detección de esas monedas como su función apropiada.) Si, por el contrario, el dispositivo de dos bits fuera enviado a Panamá por error, o si llegara por pura coincidencia, su debut no significaría nada, aunque su utilidad pudiera ser reconocida pronto —de inmediato— y apreciada por las autoridades pertinentes (aquellas que pudieron obligarla a prestar servicios en un nuevo papel) y, por consiguiente, sus estados posteriores contarían como signos de *QP*.

Presumiblemente, Fodor y otros se alegrarían de dejarme decir esto, puesto que, después de todo, el dispositivo de dos bits no es más que un artefacto. No tiene ninguna intencionalidad intrínseca, original, así que no hay ningún hecho "más profundo" del tema que podríamos tratar de descubrir. Este es sólo un asunto pragmático acerca de cuál es la mejor manera de hablar metafórica y antropomórficamente acerca de los estados del dispositivo.

Pero nos separamos cuando yo pretendo aplicar precisamente las mismas enseñanzas, las mismas normas pragmáticas de interpretación, al caso humano. En el caso de los seres humanos (por lo menos) Fodor y compañía están seguros de que esos hechos más profundos sí existen, aunque no siempre podamos encontrarlos. Es decir, que ellos suponen que, independientemente del poder que cualquier observador o intérprete pueda tener para descubrirlos, siempre hay un aspecto del asunto acerca de lo que *significa de verdad* una persona (o el estado mental de una persona). Ahora bien, podríamos llamar a su creencia compartida, una creencia de intencionalidad *intrínseca*, o hasta quizás una intencionalidad *objetiva* o *real*. Hay diferencias entre ellos acerca de cómo caracterizar y llamar a esta propiedad de la mente humana, a la que yo seguiré llamando *intencionalidad original*, pero todos están de acuerdo en que las mentes son diferentes del dispositivo de dos bits en este aspecto, y éste es el punto de desacuerdo que considero como el más fundamental entre Fodor y yo, entre Searle y yo, entre Dretske y yo, entre Burge y yo, etcétera. Una vez que estuvo afuera, muchas cosas que habían estado confundiéndome ocuparon sus lugares. Por fin entendí (y explicaré enseguida) por qué a Fodor le disgustan las hipótesis evolucionistas casi tanto como le disgusta la inteligencia artificial (véase, por ej., "Tom Swift and his Procedural Grandmother" en Fodor 1981a y el último capítulo de Fodor 1983); por qué Dretske debe llegar a extremos tan desesperados para explicar un error; por qué el "antiindividualismo" de Burge y las rumias de Kripke acerca del cumplimiento de las reglas que les impresionan a algunos filósofos como desafíos profundos y perturbadores a su complacencia, siempre me han impactado como grandes esfuerzos malgastados en tratar de tirar abajo una puerta que no está cerrada con llave.

Me separo de estos otros porque aunque pudieran coincidir conmigo (y con Millikan) sobre lo que habría que decir acerca del dispositivo de dos bits trasladado, dicen que nosotros, los seres humanos, no somos únicamente dispositivos de dos bits más imaginativos, más sofisticados. Cuando decimos que entramos en el estado de creer que estamos percibiendo una moneda de 25 centavos norteamericana (o agua auténtica como opuesta a XYZ, o una auténtica punzada de artritis) esto no es una metáfora, no es una mera manera de hablar. Un ejemplo paralelo agudizará el desacuerdo.

Supongamos que determinado ser humano, Jones, mira por la ventana y entra en el estado de creer que vio un caballo. Puede haber o no un caballo ahí afuera para que él lo vea, pero el hecho de que esté en el estado mental de creer que ve un caballo no es sólo una cuestión de interpretación (como dicen estos otros). Supongamos que el Planeta Tierra Gemelo fuera exactamente igual a la Tierra excepto por tener percaballos donde nosotros tenemos caballos. (Los percaballos parecen caballos para todos y son poco menos que indistinguibles de éstos para todos menos para los biólogos entrenados que tienen aparatos especiales, pero no son caballos, así como los delfines no son peces.) Si despachamos a Jones hacia el Planeta Tierra Gemelo, la tierra de los percaballos y lo confrontamos de manera pertinente con un percaballo, o sigue realmente en el estado de creer que ve un caballo (una creencia equivocada, no verídica) o ese percaballo lo lleva a creer, por primera vez (y verídicamente), que está viendo un percaballo. [Para bien del ejemplo supongamos que los habitantes del Planeta Tierra Gemelo llaman a los percaballos *caballos* (chevaux, Pferde, etc.), de manera que lo que Jones o un nativo del Planeta Tierra Gemelo *dice para sí* —o los demás— no cuenta para nada.] Por más difícil que sea determinar exactamente en qué estado se encuentra, está en realidad en uno u otro (o tal vez no esté verdaderamente en ninguno, tan violentamente hemos asaltado su sistema cognitivo). Cualquiera que considere que esta intuición es irresistible cree en la intencionalidad original y tiene una compañía distinguida: Fodor, Searle, Dretske, Burge y Kripke, pero también Chisholm (1956, 1957), Nagel (1979, 1986) y Popper y Eccles (1977). Cualquiera que encuentre dudosa sino totalmente descartable esta intuición puede unirse a mí, a los Churchland (véase especialmente Churchland y Churchland, 1981), Davidson, Haugeland, Millikan, Rorty, Stalnaker y nuestros distinguidos predecesores Quine y Sellars en el otro rincón (junto con Douglas Hofstadter, Marvin Minsky y casi todos los demás de *IA*).

Allí hay entonces un desacuerdo muy importante. ¿Quién tiene razón? No puedo esperar refutar la tradición opositora en la corta extensión de un capítulo pero daré dos convicciones diferentes a mi favor: demostraré con qué perplejidades se enredan Fodor, Dretzke y otros, al aferrarse a su intuición, y proporcionaré un pequeño experimento del pensamiento para motivar, si bien no para justificar, el punto de vista contrario al mío. Primero, el experimento del pensamiento.

El diseño de un robot

Supongamos que usted decidiera, por cualquier motivo, que quería experimentar la vida en el siglo veinticinco, y supongamos que la única manera conocida de mantener vivo su cuerpo durante tanto tiempo exigía que fuera puesto en una especie de dispositivo de hibernación, en el que descansaría, desacelerando y comatoso, durante todo el tiempo que usted quisiera. Usted podría hacer arreglos como para trepar a la cápsula de apoyo, ser puesto a dormir y ser despertado y liberado automáticamente en 2401. Por supuesto que éste es un tema de ciencia-ficción consagrado por el tiempo.

El diseño de la cápsula misma no es su único problema técnico, puesto que la cápsula debe estar protegida y provista de la energía suficiente (para refrigeración o lo que sea) como para más de cuatrocientos años. Naturalmente usted no podrá contar con sus hijos ni con sus nietos para dirigir esta tarea, porque habrán muerto mucho antes del año 2401, y no puede suponer que sus descendientes más lejanos, si los hubiera, se interesarían vivamente por su bienestar. De manera que debe diseñar un supersistema para proteger su cápsula y suministrarle la energía que necesita para cuatrocientos años.

He aquí dos estrategias básicas que usted podría seguir. En una, usted debería encontrar el lugar ideal, el mejor que pueda prever, para una instalación fija que estuviera bien provista de agua, luz solar y cualquier otra cosa que su cápsula (y el supersistema mismo) necesitará para todo el tiempo. El inconveniente principal de una instalación o "planta" así, es que no se la puede mover si algún daño le ocurriera por casualidad, digamos si alguien decidiera construir una autopista justo donde ella está colocada. La segunda alternativa es mucho más elaborada, pero evita este inconveniente: diseñe una instalación móvil donde alojar su cápsula junto con los sensores y dispositivos de aviso inmediato necesarios, de manera que pueda estar fuera de peligro y buscar nuevas fuentes de energía cuando las necesite. En pocas palabras, construya un robot gigante e instale la cápsula (con usted dentro) en él.

Estas dos estrategias básicas están copiadas de la naturaleza, evidentemente: corresponden aproximadamente a la división entre plantas y animales. Puesto que la segunda y más sutil estrategia, se ajusta mejor a mis propósitos, supondremos que usted decide construir un robot para alojar su cápsula. Debería tratar de diseñarla de manera que, por encima de todo lo demás "elija" las acciones diseñadas para favorecer sus mejores intereses. Los "malos" movimientos y las vueltas "equivocadas" son aquellas que tenderán a discapacitarlo para el papel de protegerlo a usted hata 2401; que es su única *raison d'être*. Este es, claramente, un problema técnico profundamente difícil, que demanda el más alto nivel de pericia para diseñar un sistema de "visión" para guiar su movilidad y otros sistemas "sensoriales" y locomotores. Y como usted estará en estado comatoso en todo respecto y, por tanto, no puede quedarse despierto para dirigir y planear sus estrategias, tendrá que diseñarlo para que genere sus propios planes en respuesta a los cambios de las circunstancias. El debe "saber" cómo "escoger" y "reconocer" y luego explotar las fuentes de energía, cómo mudarse a un territorio más seguro, cómo "prever" y después evitar los peligros. Con tanto por hacer, y hacer rápidamente, sería mejor que usted confiara, cuando pudiera, en las economías: no le proporcione a su robot más pericia que la que probablemente va a necesitar para distinguir lo que haga falta distinguir en su mundo.

El hecho de que usted no puede contar con que su robot sea el único robot así que ande por ahí con esa misión, dificultará mucho su tarea. Si su velocidad prende, su robot se encontrará compitiendo con otros (y con sus descendientes humanos) por suministros limitados de agua fresca, energía, lubricantes y demás. No hay duda de que sería prudente diseñarlo con la elaboración suficiente en sus sistemas de control como para permitirle calcular los beneficios y riesgos de cooperar con otros robots o de formar alianzas para

beneficio mutuo. (Todo cálculo así debe ser una aproximación "grosera", arbitrariamente truncada. Véase Dennett, de próxima aparición e.)

El resultado de este proyecto de diseño sería un robot capaz de exhibir autocontrol, puesto que usted le debe ceder una parte muy grande del control de su artefacto una vez que lo ponen a dormir.[2] Como tal, será capaz de derivar las propias metas subsidiarias de su evaluación de su estado presente y la importancia de ese estado para su objetivo final (que es preservarle a usted). Estos objetivos secundarios lo llevarán muy lejos en proyectos de un siglo de duración, algunos de los cuales pueden ser desatinados a pesar de los propósitos de usted. Su robot puede embarcarse en acciones antiéticas, hasta suicidas, si ha sido convencido quizá por otro robot de subordinar su propia misión de vida a otro.

Pero sin embargo, según Fodor y otros, este robot no tendría ninguna intencionalidad original sino sólo la intencionalidad que deriva de su papel de artefacto como protector suyo. Su simulacro de estados mentales no sería más que eso, no decidir ni ver ni preguntarse, ni planear *de verdad*. Sino sólo *como si* decidiera, viera, se preguntara o planeara.

Deberíamos detenernos un momento, para asegurarnos de que comprendemos lo que esta afirmación abarca. El robot imaginario es por cierto mucho más sofisticado que el modesto dispositivo de dos bits, y tal vez a lo largo del sendero hacia la mayor complicación hemos pasado de contrabando alguna nueva capacidad crucial que le otorgaría al robot nuestra clase de intencionalidad original. Observe, por ejemplo, que nuestro robot imaginario al que le hemos otorgado el poder de "planear" nuevos cursos de acción, de "aprender" de errores pasados, de formar alianzas y de "comunicarse" con sus competidores, probablemente cumpliría muy honrosamente en cualquier prueba Turing a la que lo sometiéramos (véase Dennett, 1985a). Más aun, para ejecutar todo este "planeamiento", "aprendizaje" y "comunicación", casi con certeza tendrá que estar provisto de estructuras de control ricas en poder autorreflexivo, de autocontrol, de manera que tenga un acceso de tipo humano a sus propios estados internos y sea capaz de informar, manifestar y comentar acerca de lo que "hace falta" para ser la consecuencia de sus propios estados internos. Tendrá "opiniones" acerca de lo que significan esos estados y no hay duda de que deberíamos tomar seriamente esas opiniones como muy buenas pruebas —probablemente la mejor prueba que podemos obtener fácilmente— acerca de lo que esos estados "significan" ***metafóricamente hablando*** (recuerden: no es más que un artefacto). Al dispositivo de dos bits no se le otorgó semejante capacidad de hacer vacilar nuestros juicios interpretativos mediante la manifestación de "confesiones" aparentemente secretas.

Hay varias maneras en las que uno podría responder a este experimento del pensamiento, y analizaremos las más prometedoras a su debido tiempo, pero primero quiero extraer la implicación más llamativa de mantenernos firmes en nuestra primera intuición: ningún artefacto, cualquiera que sea la can-

[2] Para más información sobre el control y el autocontrol, véase mi *Elbow Room: The Varieties of Free Will Worth Wanting* (1984), capítulo 3, "Control and Self Control"; y de próxima aparición a.

tidad de brujería que la *IA* diseñe en él, tiene otra cosa que no sea intencionalidad derivada. Si nos aferramos a este punto de vista, la conclusión que se nos impone es que nuestra propia intencionalidad es exactamente como la del robot, puesto que la historia de ciencia ficción que he contado no es nueva; no es más que una variación de la visión de Dawkins (1976) de nosotros (y todas las demás especies biológicas) como "máquinas de supervivencia" diseñados para prolongar el futuro de nuestros genes egoístas. Somos artefactos, en efecto, diseñados por encima de los eones como máquinas de supervivencia para genes que no pueden actuar rápida e informadamente en sus propios intereses. Nuestros intereses tal como los concebimos y los de nuestros genes pueden muy bien diferir, aun cuando si no fuera por los intereses de nuestros genes no existiríamos: su preservación es nuestra *raison dêtre* original aun si podemos aprender a ignorar ese objetivo e idear nuestro propio *summun bonum*, gracias a la inteligencia que nuestros genes han instalado en nosotros. ¡De manera que nuestra intencionalidad deriva de la intencionalidad de nuestros genes "egoístas"! ¡Ellos son los significadores sin significado, no nosotros!

Leyendo la mente de la Madre Naturaleza

Esta visión de las cosas, si bien proporciona una respuesta satisfactoria a la cuestión de dónde vino nuestra propia intencionalidad, parece dejarnos en realidad avergonzados, puesto que deriva nuestra propia intencionalidad de entidades —los genes— cuya intencionalidad es seguramente un caso paradigmático de mera intencionalidad *como si*. ¿Cómo podría lo literal depender de lo metafórico? Más aun, hay seguramente toda esta falta de analogía entre mi historia de ciencia ficción y la historia de Dawkins: en mi historia supuse que había una técnica consciente, deliberada, previsora, comprometida en la creación del robot, mientras que si somos, como Dawkins dice el producto de un proceso de diseño que tiene a nuestros genes como los beneficiarios principales, ese es un proceso de diseño que carece completamente de un ingeniero consciente, deliberado y previsor.

La belleza principal de la teoría de la selección natural es que nos muestra cómo eliminar a este artífice inteligente de nuestra explicación de los orígenes. Y sin embargo el proceso de selección natural es responsable de diseños de gran astucia. Es un poco excesivo concebir a los genes como diseñadores inteligentes; los genes mismos no podrían ser más estúpidos; *ellos* no pueden ni razonar ni representar o deducir nada. Ellos no hacen el diseño por sí mismos; son meramente los beneficiarios del proceso de diseño. Pero entonces, ¿quién o qué realiza el diseño? La Madre Naturaleza, por supuesto o, más literalmente, el largo y lento proceso de evolución por selección natural.

Para mí la propiedad más fascinante del proceso de evolución es su capacidad misteriosa de reflejar *algunas* propiedades de la mente humana (el Artífice inteligente), mientras está privado de otras. Si bien nunca se acentúa lo suficiente que esa selección natural opera sin ninguna visión y ningún pro-

pósito, no deberíamos perder de vista el hecho de que el proceso de selección natural ha probado ser exquisitamente sensible a las razones de ser, haciendo miriadas de "elecciones" discriminatorias y "reconociendo" y "apreciando" muchas relaciones sutiles. Para expresarlo aún más provocativamente, cuando la selección natural selecciona puede "elegir" un diseño determinado *por una razón más que por otra* sin siquiera "representar" conscientemente —¡o inconscientemente!— la elección o las razones. (Los corazones se eligieron por su excelencia como circuladores de sangre, no por el ritmo cautivante de sus latidos, aunque esa podría haber sido la razón por la que algo fue "elegido" por la selección natural.)

Entiendo que no hay ninguna representación en el proceso de selección natural y sin embargo parece ciertamente que podemos dar explicaciones principistas de características evolucionadas de diseño que invocan, en efecto, "lo que la Madre Naturaleza *pensaba*" cuando diseñó ese rasgo.[3]

Así como el licenciatario panameño de la franquicia de la Pepsi-Cola puede escoger el dispositivo de dos bits *por su talento* en reconocer balboas de 25 céntavos, lo puede adoptar *como* detector de balboas de 25 centavos, la evolución puede del mismo modo escoger un órgano por su capacidad para oxigenar la sangre y lo puede establecer *como* pulmón. Y es solamente en relación con esas "elecciones" de diseño o propósitos "endosados" por la evolución —***raison d'être***— que podemos identificar conductas, acciones, percepciones, creencias o cualquiera de las otras categorías de la psicología popular (véase Millikan, 1984-86) para una clara expresión de este punto de vista).

La idea de que somos artefactos diseñados por la selección natural es tan compulsiva como concisa; algunos llegarán tan lejos como para decir que está mucho más allá de la polémica seria. ¿Por qué, entonces, se la ataca no sólo por los creacionistas sino también (más bien subliminalmente) por personas como Fodor, Searle, Dretske, Burge y Kripke? Mi presentimiento es porque tiene dos implicaciones más bien poco obvias que algunos la encuentran terriblemente desabrida. Primero, si somos (sólo) artefactos entonces lo que nuestros pensamientos más íntimos significan —*y si significan algo*— es algo acerca de lo cual los pensadores mismos de estos pensamientos no tenemos ninguna autoridad especial. El dispositivo de dos bits se convierte en un detector de balboas sin cambiar siquiera su naturaleza interna. El estado que significaba una cosa ahora significa otra. En principio, lo mismo podría pasarnos a nosotros si no somos más que artefactos, si nuestra propia intencionalidad no es entonces original sino derivada. Aquellos que como Dretske y Burge ya han renunciado a esta doctrina tradicional de acceso privilegiado pueden descartar con un encogimiento de hombros o hasta de dar la bienvenida a esa implicación; ésa es la segunda implicación la que atacan: si somos artefactos así no sólo no tenemos ningún acceso privilegiado garantizado a

[3] "Después de todo debe de haber un número finito de principios generales que gobiernan las actividades de nuestros distintos mecanismos hacedores de estados cognitivos y usuarios de estados cognitivos y deben haber explicaciones de por qué estos principios han trabajado históricamente para ayudar a nuestra supervivencia. Suponer lo contrario es suponer que nuestra vida cognitiva es una nube epifenomental accidental que sobrevuela los mecanismos que la *evolución ideó mientras pensaba en otra cosa*. (Millikan, 1986, pág. 55, la cursiva es mía).

los hechos más profundos que fijan los signfiicados de nuestros pensamientos, sino que *esos hechos más profundos no existen*. A veces la interpretación funcional es obvia, pero cuando no lo es, cuando vamos a leer la mente de la Madre Naturaleza no hay ningún texto a interpretar. Cuando "es verdad que" la función adecuada es polémica —cuando más de una interpretación está bien sustentada— no hay verdad.

La táctica de tratar la evolución misma desde la actitud intencional necesita mayor discusión y defensa, pero quiero enfocar la tarea indirectamente. Los temas se enfocarán mejor, creo, si diagnosticamos primero la resistencia a esta táctica —y su gemela siamesa, la táctica de tratarnos a nosotros mismos como artefactos —en un trabajo reciente sobre filosofía de la mente y del lenguaje.

El error, la disyunción y la interpretación inflada

La tentativa de Dretske (1981, 1985, 1986) de tratar estos temas invoca una distinción entre lo que llama *significado natural* y *significado funcional*. El significado natural (*significado*$_n$) se define e manera tal como para excluir falsa representación lo que significa$_n$ un toque especial del timbre de la puerta depende de la integridad del circuito que causa el sonido del timbre. "Cuando hay un cortocircuito, el timbre de la puerta (independientemente de lo que estaba diseñado para indicar, independientemente de lo que indica normalmente) no indica que el botón de la puerta está siendo oprimido." "Esto es lo que *se supone* que significa$_n$, lo que estaba *diseñado para significar*$_n$ lo que (tal vez) los signos de ese tipo significa$_n$ *normalmente*, pero no lo que *sí* significa$_n$" (1986, pág. 21).

Le corresponde entonces a Dretske definir el *significado funcional*, lo que es para algo significar$_f$ que tal y tal, de manera tal como para explicar cómo un signo o estado o suceso de algún sistema puede, en ocasiones, tergiversar algo o "decir" algo falso. Pero "si estas funciones son (lo que llamaré) funciones *asignadas*, entonces el significado$_f$ está con los propósitos, intenciones y creencias de aquellos que asignan la función de la cual el significado$_f$ deriva sus poderes falsificadores" (pág. 22). Está claro que el significado del estado de aceptación *Q* del dispositivo de dos bits es justamente un significado funcional así asignado, y Dretske diría de él: "Esa es la función que le asignamos, la razón por la cual fue construido y la explicación de por qué fue construido de esa manera. Si nuestros propósitos hubieran sido otros, podrían haber significado$_f$ alguna otra cosa " (pág. 23).

Puesto que el significado funcional meramente *asignado* está "teñido", Dretske debe buscar una distinción mayor. Lo que tiene que caracterizar son las funciones *naturales* de los estados que son la contrapartida de los organismos, "las funciones que un organismo tiene, que son independientes de *nuestras* intenciones y propósitos interpretativos" (pág. 25), de manera que entonces puede definir el significado funcional natural en los términos de esas funciones.

Estamos buscando lo que *se supone* que un signo significa$_n$ donde el "se su-

pone" se paga en los términos de la función de ese signo (o sistema del signo) en la *propia* economía cognitiva del organismo (pág. 25).

El camino claro a seguir, como vimos en la última sección, es sustituir nuestras intenciones y propósitos interpretativos, las intenciones y propósitos del diseñador del organismo, la Madre Naturaleza —el proceso de selección natural— y preguntarnos qué se supone que significa o está diseñado para señalar en *este* esquema, cualquier tipo especial de señal o estado. Así como finalmente apelaríamos a las razones fundamentales de los ingenieros cuando decidiéramos acerca de la mejor explicación de la representación y la representación falsa en nuestra máquina robótica imaginaria de supervivencia, podemos apelar del mismo modo a las razones fundamentales de diseño perceptible de la selección natural al asignar contenido y, por tanto, el poder de la representación *falsa* a tipos de sucesos en los artefactos naturales —los organismos— incluidos nosotros mismos.

Pero aunque Dretske les rinde homenaje a aquellos que siguieron ese sendero evolucionista, y él mismo lo sigue con cautela a cierta distancia, ve un problema. El problema no es otro que la versión biológica de nuestra pregunta acerca de qué principio hay para decir si el estado del dispositivo de dos bits (en algún ambiente especial) significa "25 centavos aquí y ahora" o "balboas de 25 centavos aquí y ahora" o "cosa de clase *F* o de clase *G* o de clase *K* aquí y ahora". Debemos encontrar un principio interpretativo, que asigne contenido, dice Dretske, "sin hacerlo *inflando* artificialmentes las funciones naturales de estos sistemas, mientras al mismo tiempo evitamos el principio demasiado deflatorio que resuelve todo el significado funcional en el significado natural bruto donde la representación falsa es imposible.

Considérese el caso clásico de lo que el ojo de la rana le dice al cerebro de ésta (Lettvin y otros, 1959). Supongamos que provocamos a una rana para que atrape y se trague un perdigón que le arrojamos (véase Millikan, 1986). Si interpretamos la señal procedente del ojo como "diciéndole" a la rana que hay una mosca volando en su dirección, es el ojo el que le está pasando información equivocada a la rana, mientras que si interpretamos esa señal como señalando meramente una mancha movible oscura en la retina, está "diciendo la verdad" y el error se le debe asignar a alguna parte tardía del procesamiento del cerebro (véase Dennett, 1969, pág. 83). Si somos tenazmente mínimos en nuestras interpretaciones, la rana nunca comete un error puesto que todos los sucesos en el sendero pertinente de su sistema nervioso se pueden *desinterpretar* siempre agregando disyunciones (la señal significa algo menos exigente: mosca *o* perdigón *o* mancha oscura movible *o* trozo de metal de clase K *o*...) hasta que estamos de vuelta en el significado$_n$ bruto del tipo de señal, en la cual la representación falsa es imposible. Independientemente de cuántas capas de transductores contribuyen a la especificidad de una señal, siempre habrá una interpretación deflatoria de su significado como significado$_n$ a menos que relativicemos nuestra explicación a determinada presunción de la función normal (normal en el sentido de Millikan) (véase Dennett, 1969, sección 9, "Function and Contents".

Dretske está preocupado por sobredotar a los tipos de sucesos de contenido, atribuyéndoles significado más específico o complicado de lo que los

hechos dictan. Pero dada la mezquindad de la Madre Naturaleza, la ingeniera, esta sobriedad hermenéutica por otra parte laudable, lo pone a uno en peligro de no apreciar el "punto", el verdadero ingenio de sus invenciones. Una instancia especialmente instructiva de las virtudes de la interpretación funcional "inflacionista" es la respuesta especulativa de Braitenberg (1984) a la pregunta de por qué tantos seres —desde los peces a los seres humanos— están equipados con hardware de propósitos especiales que es maravillosamente sensible a los patrones visuales que exhiben simetría alrededor de un eje vertical. Cabe poca duda acerca de cuál es la descripción deflacionaria del contenido de estos intrincados transductores: señalan "un caso de simetría alrededor del eje vertical en la retina". ¿Pero por qué? ¿Para qué sirve esto? La disposición es tan común que debe tener una utilidad muy general. Braitenberg pregunta, ¿qué en el mundo (antes de que existieran las fachadas de las iglesias y los puentes colgantes) presenta una vista simétricamente vertical? Nada en el mundo de las plantas ni nada en el terreno. Sólo esto: otros animales, *¡pero solamente cuando enfrentan al observador!* (Las vistas traseras son a menudo verticales simétricamente, pero en general de manera menos notable.) En otras palabras, lo que un transductor de simetría vertical nos dice es (aproximadamente) "alguien te está mirando". Es innecesario decir que éste es un dato que bien vale la atención de un animal, puesto que el otro ser en cuyos hilos reticulares el animal comúnmente se sienta, puede ser muy bien un depredador o un rival o una pareja. De manera que no es sorprendente que el efecto normal de que el detector de simetría esté *encendido* es una reacción inmediata de la orientación y (en el caso de los peces, por ejemplo) una preparación para la huida. ¿Es inflacionario llamar a este transductor un detector de depredadores? ¿O un detector de depredadores —o parejas— o rivales? Si usted fuera contratado para diseñar un detector de depredadores de peces, ¿buscaría un transductor más infalible (pero pesado, lento) o argumentaría que ésta es en verdad la mejor clase de detector de depredadores que se puede tener, en el cual las falsas alarmas son un precio muy bajo a pagar por su velocidad y su poder para reconocer depredadores relativamente bien escondidos?

Las simetrías verticales ecológicamente insignificantes cuentan como falsas alarmas sólo si suponemos que el cableado de *propósito* especial tiene que "decirle" al organismo (aproximadamente) "alguien te está mirando". ¿Cuál es *exactamente* el contenido de esa elocución? Esta búsqueda de la precisión de la adscripción del contenido y de la independencia de la interpretación es el sello distintivo, no sólo del programa de investigación de Dretske sino de gran parte del trabajo teórico en filosofía del lenguaje y de la mente (la teoría filosófica del significado ampliamente concebido). Pero por lo menos en el caso del detector de simetría (o como quiera que queramos llamarlo) no hay ninguna respuesta de principios para eso, más allá de lo que podamos sustentar apelando a las funciones que podemos descubrir y a las que podemos encontrarles sentido de esta manera en el funcionamiento del transductor en la naturaleza.

Vimos en el caso de los artefactos de diseño humano que podríamos usar nuestra apreciación de los costes y beneficios de distintas elecciones de diseño para mejorar nuestra interpretación del talento discriminatorio del disposi-

tivo de dos bits de la mera detección del disco-de-paso-*p*-espesor-*e* y diámetro-*d* y material-*m* a la detección de la moneda de veinticinco centavos (o la detección del balboa de veinticinco centavos, según las intenciones del usuario). Esta es, si se prefiere, la táctica fundamental de la hermenéutica de los artefactos. ¿Por qué debería Dretske dejar de lado el mismo principio interpretativo en el caso del significado funcional natural? Porque a su criterio no está suficientemente basado en principios. No satisfaría nuestra ansia de una explicación de lo que *realmente* significa el hecho natural, lo que significa con el aspecto de intencionalidad "original" o "intrínseca".[4]

En *Machines and the Mental* (1985) Dretske afirma que la diferencia fundamental entre los ordenadores corrientes y nosotros es que mientras los ordenadores pueden procesar información manipulando símbolos internos de algún tipo, "no tiene acceso, por así decirlo, al *significado* de estos símbolos, a las cosas que las representaciones representan" (pág. 26). Esta manera de expresarse sugiere que Dretske está uniendo dos puntos: que algo significa, algo *para* un sistema u organismo, y que ese sistema u organismo está en posición de saber o reconocer o intuir o comprender ese hecho desde el interior.

> Salvo que estos símbolos tengan lo que podríamos llamar un significado *intrínseco* [mi énfasis], un significado que poseen que es independiente de nuestras intenciones y propósitos comunicativos, entonces este significado debe ser irrelevante a la evaluación que el mecanismo hace cuando los manipula (pág. 28).

Dretske insiste muy correctamente en que el significado que él les está buscando para los estados mentales debe *tener importancia verdadera* en y para, la vida del organismo, pero lo que él deja de ver es que el significado que busca, mientras en el caso de un organismo es independiente de *nuestras* intenciones y propósitos, no es independiente de las intenciones y propósitos de la Madre Naturaleza, y es por tanto al final tan derivado y por tanto tan sometido a la indeterminación de la interpretación, como el significado en nuestro dispositivo de dos bits.

[4] Ocurre que Dretske discute el problema de la detección de depredadores en un trozo que plantea este problema junto con su punto de vista: "Si (ciertas) bacterias no tuvieran algo dentro que significara que *eso* era la dirección del norte magnético, no podrían orientarse como para evitar el agua tóxica de superficie. Perecerían. Si, en otras palabras, los estados sensoriales internos de un animal no fueran ricos en información, significado natural intrínseco, acerca de la presencia de rapiña, depredadores, acantilados, obstáculos, agua y calor, no podría sobrevivir" (1985, pág. 29). El problema es que, dadas las exigencias conservadoras de Dretske acerca de la información, el detector de simetría no se tendría en cuenta como emisor de una señal con información (significado natural intrínseco) acerca de los depredadores, sino sólo acerca de los modelos de simetría vertical en la retina, y si bien sin duda podría ser, y sería normalmente, suplementado por más transductores diseñados para hacer distinciones de grano más fino entre los depredadores, rapiña, parejas, rivales y miembnros de especies ignorables, éstas podrían ser igualmente crudas en sus poderes discriminatorios reales. Si, como lo sugiere Dretske, algunas bacterias pueden sobrevivir sólo con detectores norte (no necesitan detectores de agua tóxica, da la casualidad) otros seres se las pueden arreglar con simples detectores de simetría, de manera que la última oración citada antes es falsa: la mayor parte de los animales sobreviven y reproducen muy bien sin el beneficio de estados que son lo bastante ricos en información (dretskeana) como para informar a sus dueños acerca de rapiñas, depredadores, acantilados, y otras cosas como ésa.

Dretske intenta escapar de esta conclusión y lograr una "determinación funcional" ante la amenazada "indeterminación funcional", ideando una historia complicada acerca de cómo el *aprendizaje* podría establecer la diferencia crucial. Según Dretske un organismo que está aprendiendo, puede, a través del proceso de exposiciones repetidas a una variedad de estímulos y el mecanismo del aprendizaje asociativo llegar a establecer un tipo de estado interno que tiene una función *definida, única* y por tanto un significado funcional.

Confrontado con nuestra máquina robótica de supervivencia imaginaria, la reacción de Dretske es suponer que con toda probabilidad algunos de sus estados sí tienen significado funcional natural (como opuesto al meramente asignado) en virtud de la historia del aprendizaje de los primeros días o años de servicio de la máquina de supervivencia. "Creo que podríamos crear (lógicamente) un artefacto que *adquiriera* su intencionalidad original, pero no uno que (por así decirlo, en el momento de la creación) *ya lo tuviera* (correspondencia personal). Las funciones soñadas y suministradas por sus ingenieros son sólo funciones *asignadas* —por más brillantemente que los ingenieros predijeran el ambiente en el que la máquina termina estando— pero una vez que el mecanismo tiene una oportunidad de responder al ambiente en un ciclo de entrenamiento o aprendizaje, sus estados tienen por lo menos la oportunidad de adquirir un significado funcional (definido, único) y no sólo el significado natural del cual está excluida la representación falsa.

No presentaré los detalles de esta ingeniosa tentativa porque, a pesar de toda su ingeniosidad, no funcionará. Fodor (1987), en el capítulo con el que empezamos, demuestra por qué. Como Fodor señala, depende primero de trazar un límite nítido entre el período de aprendizaje del organismo, cuando el estado interno está desarrollando su significado, y el período posterior cuando su significado ya está determinado. En opinión de Dretske la representación falsa es posible sólo en la segunda fase, pero cualquier límite que tracemos debe ser arbitrario. (Fodor se pregunta, ¿el silbato suena para señalar el final de la sesión de práctica y el comienzo del juego eterno?) Además, Fodor señala (no sorprendentemente) que la explicación de Dretske no puede proporcionar el significado natural funcional determinado de ningún estado representativo innato no aprendido.

Dretske no considera que esto sea un defecto. Tanto peor para los conceptos innatos, dice: "No creo que hayan o puedan haber conceptos o creencias innatos... Las creencias y deseos, las *razones* en general (la clase de cosas cubiertas por la actitud intencional), son (o así me gustaría argumentar) invocados para explicar modelos de conducta que se adquieren durante la historia de vida del organismo que exhibe la conducta (es decir, aprendido)" (correspondencia personal).

La motivación de esta actitud se puede presentar por medio de un ejemplo. Lo primero que hace un pichón de cuclillo cuando sale del cascarón, es buscar otros huevos en el nido, pues son sus potenciales competidores en la atención de sus padres adoptivos e intentar arrojarlos afuera. Con seguridad no tiene la menor noción del significado funcional de su actividad, pero ese significado está allí sin embargo —*para* el organismo— a menos que

supongamos según la última frase, que el organismo debe "tener acceso" a ese significado, tiene que estar en posición de reflejarse en él o admitirlo, por ejemplo. La razón fundamental de la actitud intencional estremecedora del cuclillo no está en cuestión; lo que queda por investigar es hasta qué punto la razón fundamental es la razón fundamental del pichón y hasta qué punto es de flotación libre: simplemente lo que Madre Naturaleza pensó (véase el capítulo 7). Para Dretske, sin embargo, ésta es una cuestión de todo o nada y está ligada a su intuición de que deben de haber significados únicos e inequívocos (funcionales naturales) para los estados mentales.

Dretske parece estar tratando de hacer dos cosas de un golpe: primero quiere trazar una diferencia principista (y de todo o nada) entre las razones fundamentales de flotación libre y —digamos— las "plenamente apreciadas"; y segundo, quiere quitar toda la inercia interpretativa en la especificación del significado "real" o "verdadero" de todos esos estados de significado apreciado. Después de todo, si apelamos a nuestras intuiciones introspectivas, eso es tal como parece: no solamente hay algo que queremos decir mediante nuestro pensamiento —algo en un todo determinado si bien a veces públicamente inefable— sino que es nuestro reconocimiento o apreciación de *ese significado*, lo que explica lo que hacemos en consecuencia. Por cierto que hay una gran diferencia entre los extremos representados por el pichón de cuclillo y, digamos, el asesino humano de cabeza y sangre frías que "sabe exactamente lo que hace y por qué", pero Dretske la quiere convertir en una diferencia incorrecta. Al hacerse eco de Searle, Dretske distinguiría en forma tajante entre sintaxis y semántica: en el asesino humano, diría: "Es este significado el que tiene la estructura (su semántica) y no sólo la estructura que tiene este significado (la sintaxis) lo que es pertinente para explicar la conducta" (correspondencia personal; véase Dretske, 1985, pág. 31). Aun si supusiéramos que Dretske pudiera motivar la colocación de semejante umbral, que divide el espectro de los casos más sofisticados entre aquellos en los que la sintaxis hace todo el trabajo y aquellos en los que la semántica entra en juego por cierto no inadvertida, está fuera de la cuestión que el rigor de una historia de aprendizaje pudiera atravesar *esa* barrera, y de alguna manera demostrarle a un organismo lo que *de verdad significaban* sus estados internos.

Además, *si* el movimiento de la historia del aprendizaje de Dretske funcionó para las representaciones aprendidas, el mismísimo movimiento podría funcionar para las representaciones innatas "aprendidas" por los antepasados del organismo por la vía de la selección natural a través de los eones. Es así, después de todo, como explicamos el advenimiento de los mecanismos innatos, surgiendo de un proceso de aproximaciones sucesivas a través del tiempo. Si, como Dretske supone, el cableado "*soft*" puede adquirir un significado funcional natural durante el lapso de vida de un organismo, gracias a sus relaciones con los hechos del entorno, el cableado "*hard*" podría adquirir el mismo significado funcional natural durante el lapso de vida de la especie.

Y de nuevo ¿cuándo hacemos sonar el silbato y congelamos, para todo el futuro, el significado de un ítem así diseñado? Lo que empezó como un dis-

positivo de dos bits se puede convertir en un dispositivo detector de balboas de 25 centavos; lo que empezó como un hueso de la muñeca puede convertirse en el dedo pulgar de un oso panda (Gould, 1980) y lo que empezó como una representación innata significando una cosa para un organismo, puede llegar a tener, con el paso del tiempo y en un ambiente nuevo un significado diferente para la progenie de ese organismo (la explicación de Dretske tiene otros problemas, algunos bien tratados por Fodor, pero pasaré por encima de ellos).

Entonces, ¿qué propone Fodor en lugar de la explicación de Dretske? El también está preocupado por la necesidad de una explicación acerca de cómo podemos imputrle el error a un organismo. ("No hay representación sin una representación falsa" sería un buen lema fodoriano). Y como Dretske, traza la distinción entre la intencionalidad original y la derivativa:

> Estoy preparado para que resulte que los anillos de humo y los de los árboles representen sólo lo relativo a nuestros intereses para predecir los incendios y cerciorarnos de la edad de los árboles, que los termostatos son sólo relativos a nuestro interés en mantener cálida la habitación, y que las palabras en inglés sólo son relativas a nuestra intención de usarlas para comunicar nuestros pensamientos. Eso significa que estoy preparado, para que sólo resulte que los estados mentales (por tanto según la TRM [la Teoría de la Representación Mental], únicamente las representaciones mentales) tienen propiedades semánticas *en primer lugar*; y que, por tanto, que una semántica naturalizada debería aplicarse, *stricto dictu*, solamente a las representaciones mentales (Fodor, 1987, pág. 99).

Y luego, como Dretske, enfrenta lo que llama el problema de la disyunción. ¿Qué fundamentos principistas u objetivos podemos tener para decir que el estado que significa "la moneda de 25 centavos aquí y ahora" (y que por tanto es un error cuando ocurre en la respuesta perceptual a un trozo de metal) en lugar de significar "moneda de 25 centavos *o* balboa de 25 centavos *o* trozo de metal de clase *K o* ... (y por lo tanto, invariablemente, no es para nada un error)? Fodor no es más inmune que Dretske (o ningún otro-a la tentación fatal de la teleología, de descubrir lo que "se supone" que el mecanismo pertinente "hace", pero él resiste en forma viril:

> No estoy seguro de que esta explicación de optimidad/teleología sea falsa pero sí la encuentro completamente insatisfactoria...Pienso que tal vez podamos lograr una teoría del error sin confiar en nociones de optimidad o teleología; y si podemos deberíamos hacerlo. Si todo lo demás es igual es seguro que cuanto menos *pop*-darwinismo haya, mejor (Fodor 1987, págs.105-6).

Reconozco el candor con el que Fodor expresa su incomodidad con apelaciones a la teoría evolutiva. (En otra parte descubre que tiene que servirse un poco de "darwinismo común" para apuntalar una explicación que necesita de las funciones de los transductores.) ¿Por qué, sin embargo, sería él tan reacio a bajar el sendero? Porque ve (me imagino) que lo más que se puede conseguir de una historia semejante, por más bien apuntalada que esté por hechos escrupulosamente reunidos del informe fósil, etc. es una historia del dispositivo de dos bits trasladado. Y Fodor quiere un significado real, origi-

nal, intrínseco —no para los estados de los artefactos ¡Dios es testigo, puesto que Searle tiene razón acerca de ellos!— a no ser por nuestras propias representaciones mentales.

¿Tiene Fodor una explicación que funcione mejor que la de Dretske? No. Es igualmente ingenioso e igualmente desdichado. Supongamos, dice Fodor: "Veo una vaca a la que estúpidamente identifico mal; la tomo, digamos, por un caballo. El tomarla así me hace el efecto de un símbolo; entonces, digo "caballo". Hay una asimetría, arguye Fodor, entre las relaciones causales que unen los "caballos" con los símbolos de los "caballos" por un lado, y las vacas con los símbolos de los "caballos", por el otro:

> En particular, confundir una vaca con un caballo no me hubiera llevado a decir caballo *excepto que había independientemente una relación semántica entre los símbolos de los caballos y el caballo*. Excepto por el hecho de que la palabra "caballo" expresa la propiedad de *ser un caballo* (es decir, que a no ser por el hecho que uno llama "caballo" *a los caballos*, no podría haber sido *esa* palabra la que al tomar una vaca por un caballo me la hubiera hecho pronunciar. Por cuanto, por contraste, puesto que el "caballo" sí significa *caballo*, el hecho de que los caballos me hagan decir "caballo" no depende de que haya una conexión semántica —o, de hecho, ninguna— entre los símbolos del "caballo" y las vacas (Fodor 1987, págs.107 y 108).

Esta doctrina de Fodor, entonces, se deletrea en los términos de contrafactuales que se mantienen bajo distintas circunstancias. Nuevamente, sin entrar en los detalles (para los cuales véase Atkins, inédito) digamos solamente que el problema es que nuestra incomodidad vuelve a surgir. ¿Cómo establece Fodor que en su idiolecto mental "caballo" significa *caballo —y no caballo u otro cuadrúpedo que se parezca a un caballo—* (o algo semejante)? O Fodor debe seguir la ruta introspectiva de Searle y decir que esto es algo que simplemente puede decir desde adentro o debe apelar a las clases mismas de las consideraciones de diseño y la "historia de optimalidad/teleología" que quiere atacar. Aquellos de nosotros a quienes siempre nos ha encantado contar esa historia sólo podemos esperar que llegue a deleitarse con ella, en especial cuando se dé cuenta de lo insípidas y difíciles de tragar que son las otras posibilidades.

Esto me lleva a Burge, quien ha construido también una serie de bombas de intuición diseñadas para revelarnos la verdad sobre el error. En una serie de trabajos Burge ha estado argumentando contra una doctrina a la que llama *individualismo*, una tesis sobre qué hechos resuelven los problemas del contenido o del significado de los estados mentales de un organismo. Según el individualismo:

> Los estados y hechos (tipos y símbolos) intencionales de un individuo, no pueden ser diferentes de lo que son dadas las historias físicas, químicas, nerviosas o funcionales del individuo, en el que estas historias están especificadas no intencionalmente de una manera que es independiente de las condiciones físicas y sociales de fuera del cuerpo del individuo (1986, pág. 4).

O en otras palabras:

> El significado o contenido de los estados internos de un individuo no podría ser diferente de lo que es, dada la historia *interna* del individuo y su constitución (considerada independiente de las condiciones exteriores a su "cuerpo").

La falsedad de esta tesis no debería sorprendernos. Después de todo, el individualismo es falso en lo que respecta a items simples como los dispositivos de dos bits. Cambiamos el significado del estado interno del dispositivo de dos bits simplemente trasladándolo a Panamá y dándole un nuevo trabajo para hacer. No se cambió nada estructural o físico dentro de él, pero el significado de uno de sus estados cambió de *Q* a *QB* en virtud de su extraña inserción en el mundo. Para atribuirles significado a los estados funcionales de un artefacto, hay que depender de las presunciones acerca de lo que se supone que hace y para obtener alguna eficacia en este sentido hay que mirar hacia el mundo más amplio de los propósitos y las proezas. La tesis antiindividualista de Burge es entonces simplemente un caso especial de una observación muy conocida: las caracterizaciones funcionales son relativas no sólo al ambiente de inserción sino también a las presunciones acerca de la optimalidad de diseño (véase, por ej., Wimsatt, 1974. Burge parece apreciar esto en la nota 18 al pie de la página 35).

Además, Burge apoya su antiindividualismo con argumentos que apelan exactamente a las consideraciones que motivaron nuestro tratamiento del dispositivo de dos bits. Por ejemplo, ofrece un argumento extendido (págs. 41 y sigs.) acerca de "una persona *P* quien normal y correctamente percibe ejemplos de una propiedad *O*" determinada, objetiva y visible, entrando en el estado *O'* y resulta que en algunas circunstancias una propiedad visible y diferente, *C*, pone a *P* en el estado *O'*. Podemos sustituir al "dispositivo de dos bits" por "*P*", "*Q*" por "*O*'", "la moneda de 25 centavos" por "*O*" y el balboa de 25 centavos" por "*C*" y notar que este argumento es nuestro viejo amigo sin agregarle ni quitarle nada.

Pero algo es diferente: Burge no deja lugar para la falta de determinación del contenido; sus formulaciones siempre presumen de que hay un hecho del asunto acerca del cual algo *precisamente* significa y aclara que tiene la intención de separarse de la escuela de implementación funcional "dependiente de la actitud". El elige "pasar por alto argumentos generalizados de que las adscripciones mentalistas son profundamente indeterminadas" (1986, pág. 6) y anuncia su realismo haciendo notar que la psicología parece presuponer la realidad de las creencias y los deseos y parece funcionar. Es decir, que la psicología hace uso de cláusulas *que* interpretadas "—o, lo que podríamos llamar con soltura 'contenido intencional'". Agrega, "no he visto ninguna razón sólida para creer que este uso es meramente heurístico, instrumentalista o de segunda clase en cualquier otro sentido" (pág. 8). Esa es la razón por la cual su tesis de antiindividualismo parece tan notable. Parece estar arguyendo en favor de la opinión notable de que la intencionalidad *intrínseca*, la intencionalidad *original* es tan sensible al contexto como la intencionalidad derivada.

Aunque Burge, como Dretske y Fodor se siente inexorablemente atraído hacia consideraciones evolucionistas, no consigue ver que su confianza en esas consideraciones mismas debe forzarlo a abandonar su poco complicado realismo acerca del contenido. Así, defiende la teoría de la visión de Marr (1982) como un ejemplo apropiadamente antiindividualista de éxito en psicología sin notar que la explicación de Marr es, como las explicaciones "técnicas" en general, dependiente de fuertes (en realidad demasiado fuertes, véase Ramachandran, 1985a, b) presunciones de optimalidad que dependen de encontrarle sentido a *lo que la Madre Naturaleza pensaba* para distintos subcomponentes del sistema visual. Sin la táctica que he estado llamando hermenéutica de los artefactos, Marr estaría privado de todo principio para la asignación de contenido. Burge mismo enuncia el resultado de la táctica:

> Los métodos de individuación y explicación están gobernados por la presunción de que el sujeto se ha adaptado al ambiente de él o de ella lo bastante como para obtener información verídica de éste en ciertas condiciones normales. Si las propiedades y relaciones que causaron normalmente impresiones visuales fueran regularmente diferentes de lo que son, el individuo obtendría información diferente y tendría experiencias visuales con un diferente contenido intencional (pág. 35).

Cuando le atribuimos contenido a un estado o estructura en el modelo de visión de Marr, debemos defender nuestra atribución alegando (en una paráfrasis de Dretske acerca del significado funcional asignado) que ésa es la función que la Madre Naturaleza le asignó a esta estructura, la razón por la cual se la construyó y la explicación de por qué fue construida de ese modo. Si sus propósitos hubieran sido otros, habría significado otra cosa.

El método que Burge apoya, entonces, debe hacer la asunción *metodológica* de que el sujeto se ha adaptado al ambiente de él o de ella lo bastante como para que cuando lleguemos a asignarles contenidos a los estados del sujeto —cuando adoptemos la actitud intencional— las atribuciones dictadas son aquellas que resultan verídicas *y útiles*. Sin la segunda condición, Burge se quedará atascado con el problema de la dispersión disyuntiva de Fodor y Dretske, puesto que siempre se puede obtener verticalidad a expensas de la utilidad agregando disyunciones. Sin embargo, la utilidad no es una propiedad determinada, objetiva, como lo aclaró el ejemplo del dispositivo de dos bits. De manera que, contrariamente a lo que Burge da por sentado, debe renunciar a la característica misma que vuelve a esta conclusión tan intrigante desde el principio: su realismo acerca del "contenido intencional", o en otras palabras su creencia de que hay una variedad de intencionalidad intrínseca u original no captada por nuestras estrategias para tratar la intencionalidad simplemente derivada como la del dispositivo de dos bits.

El realismo acerca del contenido intencional que Burge da por sentado, junto con Fodor y los demás, es presupuesto también por Putnam, cuyos experimentos del pensamiento acerca del Planeta Tierra Gemelo (Putnam, 1975a) fijó el orden del día para mucho trabajo reciente acerca de estos temas. Ahora podemos ver con claridad esto, contrastando nuestro dispositivo

de dos bits con un ejemplo putnamiano. En el caso del dispositivo de dos bits, las leyes de la naturaleza no alcanzan para elegir lo que su estado interno *significa en realidad*, so pena de hacer imposible la interpretación falsa. En relación con una representación rival u otra, varios de sus movimientos cuentan como errores, varios como representaciones falsas, pero más allá de los recursos de la hermenéutica no hay hechos más profundos como para zanjar los desacuerdos.

Téngase en cuenta entonces a los integrantes de una tribu putnamiana que tienen una palabra, digamos "glug" para el gas inflamable, explosivo que encuentran de vez en cuando en sus pantanos. Cuando los enfrentamos con cierto acetileno y lo llaman "glug", ¿están cometiendo un error o no? Podemos suponer que todo el hidrocarburo que han encontrado hasta ahora era metano, pero ellos son inexpertos acerca de la química, de manera de que no hay ningún fundamento que descubrir en su conducta pasada o disposicioens actuales que autorizarían una descripción de su estado glug como detección de metano, más que la detección de hidrocarburo gaseoso más inclusiva. Presumiblemente, el hidrocarburo gaseoso es una "clase natural" como lo son sus subespecies el acetileno, el metano, el propano y sus primos. De manera que las leyes de la naturaleza no alcanzan para priorizar una lectura por encima de la otra. ¿Existe un hecho más profundo del asunto, sin embargo, acerca de qué *quiere decir* en realidad con "glug"? Por supuesto que una vez que los eduquemos, tendrán que *llegar* a decir una cosa o la otra cuando dicen "glug", pero con anticipación a estos cambios más bien completos en sus estados cognitivos, ¿habrá ya un hecho acerca de si creen la proposición de que *hay metano presente* o la proposición de que *hay un hidrocarburo gaseoso presente* cuando se expresan diciendo "¡glug!"?

Si, como parece probable, no se puede arrancar ninguna respuesta de la explotación de la actitud intencional en el caso de ellos, yo afirmaría (junto con Quine y los otros que están de mi parte) que el significado de su creencia es simplemente indeterminado en este respecto. No se trata sólo de que yo no lo pueda decir o de que ellos no puedan: no hay nada que decir. Pero Putnam, cuando es Realista acerca del contenido intencional (véase el capítulo 10) sostendrá que hay un hecho ulterior, por más inaccesible que sea para nosotros, los intérpretes, que resuelve los problemas acerca de cuáles casos de identificación "glug" no sólo *cuentan como* errores sino que *en realidad lo son*, dado lo que "glug" significa de verdad. ¿Es este hecho más profundo algo más accesible para los nativos que para nosotros los extraños? Los Realistas se dividen con respecto a esa pregunta.

Burge y Dretske argumentan en contra de la doctrina tradicional del acceso privilegiado, y Searle y Fodor son, por lo menos, sumamente renuentes a reconocer que su pensamiento alguna vez se apoya en alguna apelación a una idea tan pasada de moda. No obstante Kripke todavía está dispuesto a revelar este secreto. En la resurrección en Kripke (1982) del enigma de Wittgenstein acerca de la obediencia a las normas encontramos a todos nuestros temas volviendo una vez más: una resistencia a la analogía con la máquina sobre la base de que el significado en los mecanismos tiene relación con "las intenciones del diseñador" (pág. 34) y el problema del error de concurrencia inmediata:

> ¿Cómo se determina cuándo ocurre un funcionamiento defectuoso?... Dependiendo del propósito del diseñador, cualquier fenómeno determinado puede o no contar con el mal funcionamiento de un mecanismo... Si algún mecanismo funciona mal alguna vez, y si de ser así cuándo, no es una propiedad del mecanismo mismo como objeto físico, sino que está bien definido en los términos de su programa únicamente, como lo estipuló su diseñador.

Esta conocida declaración acerca de la relatividad y derivacionismo del significado del mecanismo se ensambla con un desgano franco por parte de Kripke para ofrecer el mismo análisis en el caso del "mal funcionamiento" humano. ¿Por qué? Porque sugiere que nuestro significado será tan derivativo e innaccesible para nosotros directamente como para cualquier artefacto:

> La idea de que no tenemos acceso "directo" a los hechos ya sea que queramos decir plus o quus [*Q* o *QB*] en el caso del dispositivo de dos bits, es extraña en cualquier caso. ¿Acaso no sé, directamente, y con un buen grado de certeza que quiero decir *plus*?... Pueden haber algunos hechos acerca de mí a los cuales tengo un acceso indirecto, y acerca de los cuales debo formarme hipótesis tentativas: ¡pero seguramente el hecho al que me refiero por medio de *"plus"* no es uno de ellos! (pág. 40).

Esta declaración no es necesariamente Kripke hablando *in propia persona* puesto que ocurre en el medio de la respuesta dialéctica que Kripke cree que Wittgenstein le daría a un desafío particularmente escéptico, pero olvida poner alguna refutación en boca del escéptico y está dispuesto a admitir su simpatía por la posición expresada.

Y, ¿por qué no? Creo que aquí encontramos una expresión todo lo poderosa y directa como podría ser la intuición que está detrás de la creencia en la intencionalidad original. Esta es la doctrina a la que Ruth Millikan llama racionalismo del significado y es una de las grandes cargas de su importante libro *Language, Thought and Other Biological Categories*, que lo derriba de su pedestal tradicional (Millikan, 1984; véase también Millikan, inédito). En algo hay que ceder. O se debe abandonar el racionalismo del significado —la idea de ser diferente del pichón de cuclillo no sólo en tener acceso, sino también en tener acceso privilegiado a sus significados— o se debe abandonar el naturalismo que insiste en que somos, después de todo, sólo un producto de la selección natural, cuya intencionalidad es de este modo derivativa y por tanto potencialmente indeterminada.

¿Está la función en la mirada del observador?

He afirmado que las atribuciones de los estados intencionales a nosotros no se sostienen sin una apelación a las presunciones "de lo que la Madre Naturaleza pensaba", y ahora que podemos ver cuánto peso debe tener esa apelación es hora de descartar la metáfora cuidadosamente.

Algunos han percibido una contradicción o, por lo menos, una tensión insoluble, un síntoma de profunda incoherencia teórica en mi uso aparentemente obstinado de modismos antropomórficos —más específicamente in-

tencionales— para describir un proceso que, al mismo tiempo insisto en que es completamente mecánico, sin objetivos y carente de visión de futuro. Según Brentano, se supone que la intencionalidad debe ser "la señal de lo mental" y, sin embargo, la belleza principal de la teoría de Darwin es su eliminación de la mente de la explicación de los orígenes biológicos. ¡Qué propósito serio podía servir, entonces una metáfora tan flagrantemente engañosa? A Dawkins se le podría plantear el mismo desafío. ¿Cómo podría ser prudente estimular a la gente a pensar en la selección natural como si fuera un relojero, mientras se agrega que este relojero no sólo es ciego, sino que ni siquiera está tratando de hacer relojes?

Podemos ver claramente la utilidad —en realidad la inescapable utilidad— de la actitud intencional en biología viendo otros ejemplos de su aplicación. Los genes no son los únicos microagentes a los que biólogos sobrios les han otorgado poderes aparentemente cuidadosos. Tómense en cuenta los siguientes trozos de *Biochemistry* de L. Stryer (1981) citados por Alexander Rosenberg en "Intention and Action Among the Macromolecules" (1986b):

> Una tarea mucho más exigente para estas enzimas es *discriminar* entre aminoácidos similares... Sin embargo, la frecuencia observada del *error* en vivo es sólo de 1 en 3000 lo que indica que deben haber pasos de *producción* ulteriores para incrementar la fidelidad. En realidad la sintetasa *corrige* sus propios *errores*. ¿Cómo *evita* la sintetasa hidrolizar la isoleucina *AMP*, la *deseada* intermediaria? (págs. 664-5; la cursiva es de Rosenberg).

Parece evidente que ésta es una mera intencionalidad *como si*, la ficción de un teórico útil, sin duda, pero no para tomarla seria y literalmente. Las macromoléculas literalmente no evitan nada, ni desean nada, ni discriminan nada. Nosotros, los intérpretes o teóricos, les *encontramos sentido* a estos procesos dotándoles de interpretaciones mentalistas, pero (uno quisiera decir) la intencionalidad que atribuimos en estos casos no es ni verdadera intencionalidad intrínseca, ni verdadera intencionalidad derivada, sino una mera intencionalidad *como si*.

El "valor efectivo" de estas metáforas, como el valor efectivo de las metáforas acerca del egoísmo en los genes que Dawkins brinda escrupulosamente, está relativamente al alcance de la mano. Según Rosenberg, "todos los estados de una macromolécula que se pueden describir en términos cognitivos tienen una caracterización puramente física, única, de un largo manejable y una disyunción de consecuencias única, de manejo y descripción fáciles" (pág. 72) pero esto puede ser más la expresión de un ideal que los microbiólogos creen a su alcance que un *fait accompli* indiscutible. Del mismo modo nos podemos asegurar los unos a los otros que por cada máquina expendedora cuya existencia se conoce hay una explicación única, de longitud manejable, con una descripción manejable de su funcionamiento, de lo que la engañaría y por qué. Es decir que no hay detectores de monedas misteriosamente poderosos. Sin embargo, podemos identificar detectores de monedas como tales —podemos deducir que ésta es la competencia que explica su existencia— mucho antes de saber explicar, mecánicamente, cómo se logra esa competencia (o mejor aun: cómo nos aproximamos a ella).

Mientras esperamos terminar de adquirir nuestros conocimientos mecánicos, necesitamos las caracterizaciones intencionales de la biología para mantenernos informados acerca de lo que estamos tratando de explicar, y aun después de tener todas nuestras explicaciones mecánicas en su lugar, continuaremos necesitando el nivel intencional contra el cual medir las gangas que la Madre Naturaleza ha conseguido [véase Dennett, de próxima aparición].

Esta podría considerarse como la justificación metodológica suficiente para la estrategia de atribución de estados intencionales a sistemas biológicos simples, pero hay un desafío mayor a considerar. Rosenberg respalda el punto de vista —desarrollado por muchos, pero especialmente por Dennett, quien argumenta en su favor (1969 y 1983a)— de que una señal definitoria de intencionalidad significa el fracaso de la sustitución ("intensión") en los modismos que se deben utilizar para caracterizar el fenómeno. Señala luego entonces, que las atribuciones de los biólogos a las macromoléculas, los genes egoístas y cosas por el estilo no satisfacen esta condición; se puede sustituir *ad libitum* sin preocuparse por un cambio en el valor de la verdad mientras el "sujeto" (el creyente o el deseador) sea un gen o una macromolécula de algún mecanismo igualmente simple. Por ejemplo, la enzima correctora de pruebas no reconocen el error que corrige *qua error*. Y no se trata de que la sintetasa misma *desee* que la isoleucina-AMP sea el aminoácido intermediario; no tienen ninguna noción acerca de la isoleucina *qua* intermediaria.

La desaparición de la intensión en el nivel macromolecular parece al principio una objeción reveladora frente al uso persistente de los modismos intencionales para caracterizar ese nivel, pero si lo dejamos allí nos perdemos un nivel todavía más profundo en el cual reaparece la intensión faltante. La sintetasa puede no desear que la isoleucina-AMP sea el aminoácido intermediario, pero es sólo como intermediaria que la isoleucina es "deseada" para algo —como una parte insustituible de un diseño cuya razón fundamental el mismo proceso de selección natural "aprecia"—. Y si bien la enzima correctora de pruebas no tiene ninguna noción de que está corrigiendo errores *qua* errores, ¡la Madre Naturaleza sí la tiene! Es decir, que es sólo *qua* errores que los items así eliminados provocaron la creación de la competencia de "correctora de pruebas" de las enzimas en primer término. La enzima misma no es más que uno de los soldados inferiores de la naturaleza, "ellos no son los que deben razonar por qué, lo de ellos es hacer o morir", pero ***hay*** una razón por la que hacen lo que hacen, una razón reconocida por la selección natural.

¿Hay realmente alguna razón por la cual estas enzimas hacen lo que hacen? Algunos biólogos, escudriñando el abismo que se acaba de abrir, están tentados a renunciar a *toda* discusión acerca de función y propósito, y están en lo cierto acerca de una cosa; no existe ninguna posición intermedia estable.[5] Si se está listo para formular algún reclamo acerca de la función de

[5] Rosenberg (1986b): Entre los biólogos evolucionistas están aquellos que condenan la identificación de las estructuras anatómicas como poseedoras de significación de adapción específica, sobre la base de que esas estructuras no enfrentan la selección individualmente, sino sólo en la compañía del resto del organismo. Esto hace indeterminadas a las adscripciones de "contenido" a una parte del organismo, puesto que una adscripción diferente junto con otros ajustes en nuestras

las entidades biológicas —por ejemplo, si se desea mantener que está listo para formular algún reclamo acerca de la función de las entidades biológicas —por ejemplo, si se desea mantener que es perfectamente respetable decir que los ojos son para ver y las alas del águila para volar— se adquiere un compromiso con el principio de que la *selección* natural tiene bien puesto el nombre. En los términos de Sober (1984) no sólo hay selección *de* características sino también selección *para* las características. Si usted pasa a aseverar tales reclamos, descubre que resisten la sustitución a la manera clásica de los contextos intencionales. Así como el rey Jorge IV se preguntaba si Scott era el autor de *Waverley* sin preguntarse si Scott era Scott, así la selección natural "deseó" que la isoleucina fuera la intermediaria sin desear que la isoleucina fuera isoleucina. Y sin esta habilidad "discriminatoria" de la selección natural, sostener las interpretaciones funcionales sería imposible.

Por cierto que podemos describir todos los procesos de la selección natural sin apelar a ese lenguaje intencional, pero a un costo enorme de pesadez, falta de generalidad y detalles no deseados. Nos perderíamos el modelo que estaba allí, el modelo que permite la predicción y apoya lo contrafactual. La pregunta "¿por qué?" que podemos formular acerca de la ingeniería de nuestro robot, que tiene respuestas que aluden a los razonamientos conscientes, deliberados, explícitos de los ingenieros (en la mayor parte de los casos) tienen sus paralelos cuando el tema son los organismos y su "ingeniería". Si calculamos las razones fundamentales de estos trozos de genio orgánico nos quedaremos teniendo que atribuir —pero no de ninguna manera misteriosa— una apreciación o reconocimiento emergente de esas razones fundamentales a la selección natural misma.

¿Cómo puede la selección natural hacer esto sin inteligencia? No busca conscientemente estas razones fundamentales, pero cuando tropieza accidentalmente con ellas, las exigencias brutas de la réplica aseguran que "reconoce" su valor. Se crea la ilusión de inteligencia a causa de nuestra perspectiva limitada del proceso; la evolución puede muy bien haber probado todos los "movimientos estúpidos" además de los "movimientos inteligentes", pero los movimientos estúpidos, al ser fracasos, desaparecieron de la vista. Todo lo que vemos es la sucesión ininterrumpida de triunfos.[6] Cuando nos asignamos la

---

identificaciones de adaptación pueden resultar en el mismo nivel de aptitud para el total del organismo. En la filosofía de la psicología el dual de esta tesis se refleja en la indeterminación de la interpretación.

[6] Esta ilusión tiene la misma explicación que la ilusión explotada por los artistas embaucadores en la "pirámide revendedora" (Dennett, 1984d, págs. 92 y siguientes). Schull (de próxima aparición) arguye que el proceso de selección natural no siempre necesita ser *perfectamente* estúpido, una prueba de errores por aproximación sucesiva de la fuerza bruta de todas las posibilidades. Gracias al efecto Baldwin, por ejemplo, se puede decir que las especies mismas hacen una prueba preliminar de algunas de las posibilidades del espacio fenotípico, permitiendo así una exploración más eficiente, por medio de las genomas, de todo el espacio del paisaje adaptativo. Tal como los seres que pueden "probar opciones en sus mentes" antes de comprometerse con la acción, son más sagaces que esos seres simplemente skinnerianos que sólo pueden aprender por las pruebas de aproximaciones sucesivas del mundo real (Dennett, 1974a), así las especies que "prueban opciones en su plasticidad fenotípica" pueden —sin ninguna magia lamarckiana— darle una mano a la Madre Naturaleza en su propio rediseño.

tarea de explicar por qué *ésos* fueron los triunfos, descubrimos la razón de las cosas, las razones ya "reconocidas" por el éxito relativo de los organismos dotados de esas cosas.

Las razones originales y las respuestas originales que las "descubrieron" no fueron nuestras ni de nuestros antepasados mamíferos, sino de la naturaleza. La naturaleza apreció estas razones sin representarlas.[7] Y el proceso de diseño mismo es la fuente de nuestra propia intencionalidad. Nosotros, los representantes de la razón, los autorrepresentantes somos un producto tardío y especializado. Lo que esta representación de nuestras razones nos proporciona es la previsión: el poder anticipador del tiempo real del que carece totalmente la Madre Naturaleza. Como producto tardío y especializado, un triunfo de la alta tecnología de la Madre Naturaleza, nuestra intencionalidad es altamente derivada, y exactamente en el mismo modo en que está derivada la intencionalidad de nuestro robot (y hasta la de nuestros libros y mapas). Una lista de compras que se tiene en la mente no tiene más intencionalidad intrínseca que una lista de compras escrita en un papel. Lo que significan los items de la lista (si es que significan algo) queda establecido por el papel que juegan en el esquema mayor de los propósitos. Podemos llamar real a nuestra propia intencionalidad, pero debemos reconocer que está derivada de la intencionalidad de la selección natural, que es igualmente real, pero menos fácilmente discernible debido a la vasta diferencia en la escala de tiempo y en el tamaño.

De manera que si ha de haber alguna intencionalidad original —original sólo en el sentido de no derivar de ninguna otra fuente ulterior— la intencionalidad de la selección natural se merece el honor. Lo particularmente satisfactorio acerca de esto, es que terminamos el amenazado retroceso de la derivación con algo de la clase metafísica correcta: una fuente *cierta* y *no representativa* de nuestros propios poderes videntes y no videntes de representación. Como Millikan (inédito, manusc., pág. 8) dice, "aquí la *raíz* de los propósitos deben ser los propósitos inexpresados".

Esto resuelve el problema de la regresión sólo planteando lo que seguirá pareciendo ser un problema para cualquiera que crea todavía en la intencionalidad intrínseca, determinada. Puesto que al principio *no* fue el Verbo no hay ningún texto que se pudiera consultar para resolver las cuestiones no resueltas acerca de la función y por lo tanto acerca del significado. Pero recordemos; la idea de que una palabra —hasta un Verbo— *pudiera* tener así un significado oculto que resolviera ese problema es en sí misma un callejón sin salida.

Hay sólo una ilusión poderosa más que explorar. Creemos tener un buen modelo de función *determinada*, incontrovertible porque tenemos casos de diseño consciente, deliberado cuya historia conocemos con todos los detalles

[7] Si continuamos la extensión de la aplicación de la actitud intencional a las especies de Schull (de próxima aparición) vemos que en un sentido hay representación en el proceso de selección natural, después de todo, en la historia de la proliferación variable de las "expresiones" fenotípicas de las ideas genotípicas. Por ejemplo, podríamos decir acerca de una serie determinada que varios de los integrantes de su subpoblación habían "evaluado" opciones especiales de diseño y vuelto al pozo de genes de la especie con sus veredictos, algunos de los cuales fueron aceptados por la especie.

que usted quiera. *Conocemos* la *raison d'être* de un reloj de bolsillo, o de una gallina ponedora, porque la gente que los diseñó (o rediseñó) nos ha dicho, con palabras que entendemos, exactamente lo que tenían en la mente. Es importante reconocer, sin embargo, que por más incontrovertibles que puedan ser estos hechos históricos, sus proyecciones en el futuro no tienen ninguna significación garantizada. Alguien podría arrancar con el objetivo más ferviente, articulado y claro de fabricar un reloj de bolsillo y lograr fabricar algo que era un reloj de bolsillo terrible, inútil, o un pisapapeles soberbio, por pura casualidad. ¿Cuál de las dos es? Siempre es posible insistir en que una cosa es, esencialmente, lo que su creador se propuso que fuera, y luego cuando los hechos históricos dejan pocas dudas acerca de ese hecho psicológico, la identidad de la cosa está más allá de la cuestión. En la crítica literaria, a esa insistencia se la conoce, tendenciosa pero tradicionalmente, como la falacia intencional. Hace mucho tiempo que se discutió en esos círculos que uno no resuelve ninguna cuestión del significado de un texto (u otra creación artística) "preguntándole al autor". Si uno pone a un lado al autor, el creador original, como una guía definitiva y privilegiada del significado, puede suponer que los lectores posteriores (usuarios, seleccionadores) son indicadores igualmente importantes de "el" significado de algo, pero por supuesto que son igualmente falibles —si sus respaldos se toman como pronosticadores de significación *futura*— y por otra parte sus respaldos son sólo más hechos históricos inertes. De manera que hasta el papel del licenciatario de la franquicia de la Pepsi-Cola al elegir el dispositivo de dos bits *como* un detector de balboas de 25 centavos es sólo un acontecimiento más en la historia de vida del dispositivo, tan necesitado de interpretación como cualquier otro, puesto que este contratista puede ser un tonto. Curiosamente, entonces, tenemos *mejores* fundamentos para hacer atribuciones funcionales fiables (atribuciones funcionales que probablemente continuarán siendo valiosos auxiliares de la interpretación en el futuro) cuando pasamos por alto "lo que dice la gente" y leemos la función que podemos entre las hazañas discernibles de los objetos en cuestión más que de la historia del desarrollo del diseño.

No podemos empezar a encontrarles sentido a las atribuciones funcionales hasta que no abandonamos la idea de que tiene que haber una respuesta *correcta*, determinada para la pregunta: ¿para qué sirve? Y si no hay ningún hecho más profundo que pudiera resolver esa cuestión, no puede haber ningún hecho más profundo que resuelva su gemela: ¿qué significa?[8]

Los filósofos no están solos en su inquietud con las apelaciones a la optimidad de diseño y a lo que la Madre Naturaleza debe de haber tenido en mente. El debate en la biología entre los adaptacionistas y sus críticos es un frente diferente de la misma guerra de nervios (véase el capítulo 7). La afinidad de los temas surge con toda claridad, tal vez, en las reflexiones de

[8] La tesis de Quine acerca de la indeterminación de la traducción radical es entonces consistente con su ataque al esencialismo: si las cosas tuvieran esencias intrínsecas reales, podrían tener significados intrínsecos reales. Los filósofos se han inclinado a encontrar el escepticismo de Quine acerca de los significados fundamentales mucho menos creíble que sus animadversiones hacia las esencias fundamentales, pero eso sólo demuestra el dominio insidioso del racionalismo del significado sobre los filósofos.

Stephen Jay Gould sobre el pulgar del oso panda. Un tema central de la teoría evolucionista desde Darwin hasta el presente (especialmente en los escritos de François Jacob (1977) sobre el *bricolage* o "chapucería" de los procesos de diseño evolucionistas y en aquellos de Gould mismo) es que la Madre Naturaleza es una satisfacedora, una hacedora oportunista, no un "ingeniero ideal" (Gould, 1980, pág. 20). El célebre pulgar del oso panda "no es, anatómicamente, en absoluto un dedo" (pág. 22) sino un hueso sesamoide de la muñeca arrancando de su primer papel y puesto a servir forzosamente (por la vía de algún rediseño) *como* un pulgar. "El pulgar sesamoide no gana ningún premio en ninguna competencia de ingeniería... pero realiza su trabajo" (pág. 24). Es decir, que hace su trabajo *en forma excelente* y así es como podemos estar tan seguros de cuál es su trabajo; es obvio para lo que sirve este apéndice. ¿De manera que es como el detector de balboas de 25 centavos que empezó su vida como un dispositivo de dos bits? Gould cita a Darwin:

> Aunque un órgano pueda no haber sido formado originalmente para algún propósito especial, si ahora sirve para este fin estamos justificados al decir que está especialmente ideado para él. Según el mismo principio, si un hombre fabricara una máquina para algún uso especial, pero usara ruedas, resortes y poleas viejas sólo ligeramente reformadas, se podría decir que toda la máquina con todas sus partes estaba especialmente ideada para este fin. De este modo, a lo largo de toda la naturaleza casi todas las partes de cada ser viviente probablemente ya han servido en un estado ligeramente modificado para distintos propósitos y han actuado en la maquinaria viviente de muchas formas específicas, antiguas y diferentes.

"Podemos no sentirnos halagados", sigue diciendo Gould, "por la metáfora de las ruedas y poleas restauradas, pero téngase en cuenta lo bien que funcionamos" (pág. 26). Según este trozo parecería que Gould era un partidario sin problemas de la metodología de leer la función de la hazaña, que es por cierto lo que Darwin está defendiendo. Pero en realidad, Gould es un crítico muy bien conocido del pensamiento adaptacionista, que encuentra una "paradoja" (pág. 20) en esta mezcolanza de chapucería y teleología. No hay ninguna paradoja, hay sólo la "indeterminación funcional" que Dretske y Fodor ven y rehúyen. La Madre Naturaleza no se compromete explícita y objetivamente con ninguna atribución funcional; todas esas atribuciones dependen del conjunto mental de la actitud intencional en la que damos por sentada lo óptimo para interpretar lo que encontramos. El pulgar del oso panda no fue *en verdad* más un hueso de la muñeca de lo que es un pulgar. No es probable que nos sintamos desconcertados, en nuestra interpretación si lo consideramos *como* un pulgar, pero eso es lo mejor que podemos decir aquí o en cualquier otra parte.[9]

[9] Podemos completar nuestro recorrido de los ejemplos del dispositivo de dos bits en la literatura considerando la discusión de Sober (1984) del molesto problema de si llamar a *las primerísimas* aletas dorsales que aparecieron en un estegosauro una adaptación *para el enfriamiento*.

Supongamos que el animal tenía el rasgo a causa de una mutación, más que por la selección. ¿Podemos decir que el rasgo era una adaptación *en el caso de ese solo organismo?* He aquí algunas opciones: 1) Aplique el concepto de adaptación a poblaciones históricamente persistentes, no a organismos únicos; 2) admita que las aletas dorsales fueron una adaptación del organismo origi-

Después de todos estos años, apenas estamos poniéndonos de acuerdo con esta implicación inquietante de la destrucción por parte de Darwin del argumento del diseño: no hay ningún manual fundamental del usuario en que estén oficialmente representados las *verdaderas* funciones y los *verdaderos* significados de los artefactos biológicos. No hay más fundamentos sólidos para lo que podríamos llamar funcionalidad original que la que hay para su descendiente cognitivista, la intencionalidad original. No puede haber realismo acerca de los significados sin realismo acerca de las funciones. Como lo señala Gould, "podemos no sentirnos halagados" —especialmente cuando aplicamos la moraleja a nuestro sentido de nuestra propia autoridad acerca de los significados, pero no tenemos ninguna razón para no creer en ella.

---

nal de acuerdo con lo que ocurrió después; 3) niegue que las aletas dorsales sean adaptaciones para el organismo inicial pero diga que son adaptaciones cuando ocurren en organismos posteriores. Me inclino a preferir la elección 3 (pág. 197).

Véase también su discusión acerca del significado funcional del grosor de la piel de la *drosofila* trasladada a ambientes distintos (págs. 209-10) y su discusión (pág. 306) de cómo se podría calcular qué propiedades están siendo escogidas *para* por la Madre Naturaleza (ahora con el aspecto del entrenador de los remeros de Dawkins): "¿El entrenador escogía combinaciones de remeros? ¿Escogía determinados remeros? No nos hace falta psicoanalizar al entrenador para descubrirlo. No es el psicoanálisis, pero por lo menos la adopción de la actitud intencional nos ayudará a realizar la mecánica inversa que necesitamos hacer para conseguir respuestas a estas preguntas.

9

El pensamiento veloz*

Una última vez más consideramos el argumento del cuarto chino de John Searle (Searle, 1980, y de próxima aparición). Este argumento da a entender la futilidad de la "IA fuerte", el punto de vista de que el "ordenador digital adecuadamente programado con las entradas y salidas de datos correctas tendría, por eso, una mente exactamente en el mismo sentido en el que la tienen los seres humanos (Searle, de próxima aparición). El insiste en que su argumento es "muy simple"; no entiende que sólo un tonto o un fanático podría no ser persuadido por él. [1]

Creo que sería provechoso acercarse a estas afirmaciones debatidas con frecuencia desde un punto de vista sustancialmente diferente. No tiene sentido repasar una vez más la historia del cuarto chino y los diagnósticos que compiten acerca de lo que está ocurriendo en su interior (el no iniciado puede encontrar el artículo original de Searle, vuelto a imprimir en su totalidad, seguido por lo que todavía es el diagnóstico definitivo de su funcionamiento, en Hofstadter y Dennett, 1981, págs. 353-82). (El cuarto chino no es él mismo el argumento, en todo caso, sino más bien un impulsor de la intuición, como lo reconoce Searle: "El sentido de la parábola del cuarto chino es simplemente recordarnos la verdad de este punto más bien obvio: el hombre que está en el cuarto tiene toda la sintaxis que le podemos brindar, pero no adquiere por medio de ella, la semántica pertinente" (Searle, de próxima aparición).

He aquí el simple argumento de Searle, al pie de la letra:

> Proposición 1: Los programas son puramente formales (es decir, sintácticos).
> Proposición 2: La sintaxis no es equivalente a la semántica ni es suficiente por sí misma.
> Proposición 3: Las mentes tienen contenidos mentales (es decir, contenidos semánticos).

* Las primeras versiones de las ideas de este capítulo aparecieron en "The Role of the Computer Metaphor in Understanding the Mind" (1984e) y se extraen trozos de "The Myth of Original Intentionality" en W. Newton Smith y R. Viale, comps., ***Modelling the Mind*** (Oxford, Oxford University Press, de próxima aparición) y vuelto a imprimir con permiso.

[1] "Ya no se puede dudar más de que la concepción clásica de la IA, el punto de vista que he llamado IA fuerte, es evidentemente muy falsa y se apoya en errores muy simples" (Searle, de próxima aparición, manusc. pág. 5).

Conclusión 1: Tener un programa —cualquier programa por sí mismo— no es suficiente ni equivalente a tener una mente.

Searle desafía a sus opositores a demostrar explícitamente qué creen que está mal en el argumento, y eso es exactamente lo que haré concentrándome primero en la conclusión, la que, a pesar de su simplicidad y llaneza aparentes, es sutilmente ambigua. Empiezo por la conclusión porque me he enterado de que muchos de los seguidores de Searle están mucho más seguros de su conclusión que del camino por el que llega a ella, de manera que tienden a considerar las críticas a los pasos que siguió como simples sutilezas académicas. Una vez que hemos visto qué está mal en la conclusión podemos volver atrás para diagnosticar los pasos en falso que condujeron a Searle allí.

¿Por qué están algunas personas tan seguras de la conclusión? Supongo que tal vez, en parte, porque desean con vehemencia que sea verdad. (Uno de los pocos aspectos del prolongado debate acerca del experimento de Searle que me ha fascinado es la intensidad del sentimiento con el que muchos —los legos, los científicos, los filósofos —adoptan la conclusión de Searle.) Pero tal vez también porque la están confundiendo con un vecino cercano mucho más defendible con quien se la confunde con facilidad. Se podría suponer muy bien que las dos proposiciones siguientes llegaban prácticamente a lo mismo.

(S) Ningún programa de un ordenador podría ser nunca suficiente por sí mismo pra producir lo que un cerebro humano orgánico, con sus poderes causales especiales, puede producir de manera demostrable: fenómenos mentales con contenido intencional.

(D) No hay forma de programar un ordenador digital electrónico de manera que pueda producir lo que un cerebro humano orgánico, con sus poderes causales especiales, puede producir de manera demostrable: el control de la actividad intencional rápida, inteligente, que exhiben los seres humanos normales.

Como lo sugieren las iniciales, Searle ha aprobado la proposición (S) como una versión de su conclusión, mientras que yo estoy por presentar un argumento en favor de la proposición (D) que será marcadamente distinta de la versión de Searle únicamente después de ver cómo continúa la discusión. Pienso que la proposición (S), dado lo que Searle quiere decir con ella, es incoherente, por razones que explicaré a su debido tiempo. No estoy convencido de que la proposición (D) sea verdadera, pero considero que es una afirmación empírica coherente en favor de la cual hay algo interesante que decir. Además estoy seguro de que (D) no es para nada lo que Searle afirma en (S) —y esto me lo ha confirmado Searle en la correspondencia personal— y que mi defensa de (D) es congruente con mi defensa de la IA fuerte.

De manera que todo el que piense que ningún creyente en una IA fuerte podría aceptar (D) o que piense que (S) y (D) son equivalentes o que piense que (S) deriva de (D) o viceversa, debería interesarse en ver cómo se puede argumentar en favor de una sin la otra. La diferencia crucial es que, aunque tanto Searle como yo estamos impresionados por los poderes causales de la mente humana, estamos completamente en desacuerdo acerca de qué poderes causales importan y por qué. De manera que mi trabajo es

fundamentalmente aislar los supuestos poderes causales del cerebro según Searle y demostrar lo extraños —lo esencialmente incoherentes— que son.

Tenemos que aclarar primero una confusión menor acerca de lo que Searle quiere decir con "el programa de un ordenador por sí mismo". Hay un sentido en el que es perfectamente evidente que ningún programa de un ordenador puede *por sí mismo* producir alguno de los efectos mencionados en (S) y (D): ningún programa de computación que está guardado sin uso en un estante, una simple secuencia abstracta de símbolos, puede causar nada. Por sí mismo (en este sentido), ningún programa de un ordenador puede siquiera sumar 2 más 2 y obtener 4; en este sentido, ningún programa de un ordenador puede hacer que se produzca el procesamiento de palabras, ni mucho menos producir fenómenos mentales con contenido intencional.

Tal vez parte de la convicción que Searle ha generado en el sentido de que es *sencillamente obvio* que ningún programa de computación podría "por sí mismo" "producir intencionalidad" deriva en realidad de confundir esta afirmación obvia (e irrelevante) con algo más sustantivo y dudoso: que ningún programa corriente de un ordenador al no ser utilizado en forma concreta, podría "producir intencionalidad". Pero solamente la segunda afirmación es un desafío para la IA, de manera que demos por sentado que Searle, por lo menos, no está en absoluto confundido a este respecto y cree que ha demostrado que ninguna encarnación material corriente de un programa "formal" de un ordenador podría "producir intencionalidad" o ser capaz de "causar fenómenos mentales" (Searle, 1982) puramente en virtud de ser una encarnación de ese programa formal.

El punto de vista de Searle entonces llega a esto: tome un objeto material (cualquier objeto material) que *no* tenga el poder de causar fenómenos mentales. Usted no lo puede convertir en un objeto que sí tenga el poder de producir fenómenos mentales mediante el simple recurso de programar —reorganizando las dependencias condicionales de las transiciones entre los estados— los poderes causales cruciales del cerebro no tienen nada que ver con los programas que se podría decir que dirigen, de manera de que "darle a algo el programa correcto" no sería un modo de darle una mente.

Por el contrario, mi opinión es que, esa programación, ese rediseño de las regularidades de transición del estado de un objeto, es precisamente lo que podría proporcionarle a algo una mente (en el único sentido en el que tiene sentido pero que en realidad es empíricamente poco posible que las clases apropiadas de programas ¡puedan andar basados en otra cosa que no sea el cerebro humano orgánico! Para saber por qué esto podría ser así, consideremos una serie de argumentos inconvincentes, cada uno de los cuales cede terreno (mientras extractamos un trozo).

La divertida fantasía de Edwin A. Abbott *Flatland: a Romance in Many Dimensions* (1984) cuenta la historia de seres inteligentes que viven en un mundo bidimensional. Algún aguafiestas, cuyo nombre afortunadamente he olvidado, objetó una vez que la historia de Flatland no podía ser verdad (¿quién pensó jamás lo contrario?) porque no podía haber un ser inteligente en nada más que dos dimensiones. Para ser inteligente, argüía este escéptico, se necesita un cerebro ricamente interconectado (o sistema nervioso, o *cierta*

clase de sistema de control complejo, muy interconectado) y en sólo dos dimensiones no se pueden unir por medio de cables nada más que cinco cosas una contra la otra: por lo menos un cable se debe cruzar con otro cable, lo que exigiría una tercera dimensión.

Esto es posible, pero falso. John von Neumann probó hace muchos años que se podía producir una máquina de Turing en dos dimensiones y Conway construyó de verdad una máquina de Turing universal en su mundo de vida bidimensional. Los cruzamientos son de verdad deseables, pero hay varias maneras de prescindir de ellos en un ordenador o en un cerebro (Dewdney, 1984). Por ejemplo, está la manera en que los cruzamientos a menudo se eliminan en los sistemas de autopistas: mediante intersecciones de "semáforo" donde fragmentos de información (o lo que sea) puede turnarse para cruzar las trayectorias mutuas. El precio que se paga, aquí como en la autopista, es la velocidad de la transacción. Pero *en principio* (es decir, si el tiempo no fuera un objetivo) una entidad con un sistema nervioso todo lo interconectado que se quiera se puede realizar en dos dimensiones.

No obstante, la velocidad es "indispensable" para la inteligencia. Si no se pueden calcular los fragmentos pertinentes del entorno cambiante lo bastante rápidamente como para valerse por uno mismo, no se es *prácticamente* inteligente, por más complejo que se sea. Por supuesto que todo esto demuestra que la velocidad *relativa* es importante. En un universo en el que los acontecimientos ambientales que importaban se desplegaban cien veces más despacio, una inteligencia podía funcionar más lentamente, por el mismo factor, sin ninguna pérdida; pero traída de vuelta a nuestro universo sería más que retardada. (¿Las víctimas de la enfermedad de Parkinson que tienen "hipotensión ostostática" son dementes o están sus cerebros —como lo han sugerido algunos— sólo terriblemente retardados, pero normales en cualquier otro sentido? Ese estado es, sin embargo, discapacitante aunque sea "simplemente" un cambio de ritmo.)

Es así como no es ningún accidente que nuestros cerebros utilicen las tres dimensiones espaciales. Esto nos proporciona una conclusión empírica modesta pero bien sustentada: nada que no fuera tridimensional podría controlar la actividad intencional veloz e inteligente que exhiban los seres humanos normales.

Los ordenadores digitales son tridimensionales, pero casi todos ellos son fundamentalmente *lineales* en cierto modo. Son las máquinas de von Neumann: seriadas, no paralelas en su construcción, y así capaces de hacer solamente una cosa a la vez. Se ha convertido en un lugar común en estos días que aunque una máquina von Neumann, como la máquina de Turing universal de la que desciende, pueda *en principio* computar cualquier cosa que cualquier ordenador puede computar, muchos cálculos interesantes —especialmente en áreas cognitivas tan importantes como el reconocimiento de modelos y las búsquedas en la memoria— no pueden ser efectuados por ellos en lapsos razonablemente cortos, aun con el hardware corriendo al límite absoluto de velocidad: la velocidad de la luz, con distancias microscópicas entre los elementos. La única manera de efectuar estos cálculos en lapsos realistas de tiempo verdadero es usar hardware de procesamiento masiva-

mente paralelo. Esa es, en verdad, la razón por la cual ese hardware está siendo diseñado y construido ahora en muchos laboratorios de IA.

No es ninguna novedad que el cerebro proporciona todas las pruebas posibles de que tiene una arquitectura paralela masiva —de millones si no de miles de millones de canales de amplitud, todos capaces de actividad simultánea. Esto tampoco es accidental, es de presumir. De manera que los poderes causales necesarios para controlar la actividad intencional veloz e inteligente exhibida por los seres humanos normales sólo se puede lograr en una procesadora paralela masiva, tal como el cerebro humano (obsérvese que no he intentado una prueba *a priori* de esto, me conformo con aceptar la probabilidad científica.) Puede muy bien parecer, sin embargo, que no hay ninguna razón por la cual la procesadora paralela masiva de alguien, tenga que estar hecha de material orgánico. En realidad, las velocidades de transmisión de los sistemas electrónicos son órdenes de una magnitud mayor que la de las velocidades de transmisión en las fibras nerviosas, de manera que un sistema electrónico paralelo podría ser miles de veces más veloz (y más fiable) que cualquier sistema orgánico. Tal vez, pero de nuevo tal vez no.

Sejnowski (de próxima aparición), quien calcula la tasa promedio de procesamiento del cerebro a 10^{15} operaciones por segundo lo que es *cinco órdenes de magnitud* más veloz que hasta el producto corriente de las procesadoras electrónicas paralelas masivas; brinda una discusión esclarecedora de la velocidad relativa de los cerebros humanos y los ordenadores de hardware existentes y proyectados. Puesto que la relación de la velocidad de la computación con el coste ha decrecido en unas cinco veces en los últimos treinta y cinco años, se podría extrapolar que en pocas décadas tendremos hardware accesible, construible, que pueda equipararse al cerebro en velocidad, pero Sejnowski conjetura que esta meta no se puede alcanzar con la tecnología electrónica existente. Cree que quizás un cambio hacia la computación óptica proporcionará el punto de penetración.

Aun cuando la computación óptica pudiera proporcionar 10^{15} operaciones por segundo a un coste razonable, eso no sería una velocidad suficiente ni siquiera aproximada. Los cálculos de Sejnowski podrían probar que están subestimando los requisitos por orden de magnitud si el cerebro utiliza al maximo sus materiales. *Podríamos* necesitar pasar a la computación *orgánica* para obtener la velocidad necesaria. (Es aquí donde el argumento en favor de la proposición (D) se vuelve muy especulativo y polémico.) Supongamos —y esto no es muy probable pero apenas refutable— que la destreza procesadora de información de cualquier neurona única (su función de entrada y salida pertinente) depende de características o actividades en las moléculas orgánicas subcelulares. Es decir, supongamos que esa información procesada en el nivel de la enzima (digamos) jugara un papel crítico en modular la computación o procesamiento de información de las neuronas individuales —cada neurona, un ordenador diminuto que usa sus entrañas macromoleculares para computar su compleja e intensiva función de entrada y salida. Entonces en realidad podría no ser posible armar un modelo o simulación de la conducta de una neurona que pudiera duplicar las hazañas de procesamiento de información de la neurona *o el tiempo real.*

Esto se debería a que su modelo de ordenador necesitaría verdaderamente ser diminuto y veloz, pero no tan diminuto (y, por tanto, no tan veloz) como las moléculas individuales que se están copiando. Aun con la ventaja en velocidad de la electrónica (o la óptica) sobre la transmisión electroquímica en las ramificaciones axonales podría así resultar que los microchips son incapaces de avanzar al mismo paso que las operaciones intracelulares neuronales en la tarea de determinar exactamente cómo regular estas ramificaciones de salida lentas y pesadas.

Jacques Monod, quien habla del poder "cibernético (es decir, telenómico) a disposición de una célula equipada con cientos o miles de estas entidades microscópicas, todas mucho más inteligentes que el demonio de Maxwell-Szilard-Brillouin" presenta una versión de esta idea (Monod, 1971, pág. 69). Es una idea intrigante, pero por otra parte la complejidad de la actividad molecular en los cuerpos celulares de las neuronas puede muy bien tener sólo una significación local, no conectada con las tareas de procesamiento de información de las neuronas, en cuyo caso el tema decae.

Algunos piensan que hay fundamentos más decisivos para descartar la posibilidad de Monod. Rodolfo Llinas me ha afirmado en una conversación que no hay forma de que una neurona aproveche la velocidad del rayo y el poder "cibernético" de sus moléculas. Aunque las moléculas *individuales* puedan realizar un procesamiento veloz de la información, no se las puede hacer propagar y ampliar rápidamente estos efectos. Los hechos de ramificación en los axones neurales que tendrían que regular, son órdenes de magnitud mayor y más poderosa que sus propios cambios de estado de "salida", y el lento proceso de difusión y amplificación de su "señal" malgastaría todo el tiempo ganado mediante la miniaturización. Otros neurocientíficos con los que he conversado se han mostrado menos confiados en que la difusión relativamente lenta sea el único mecanismo disponible para la comunicación intracelular, pero no han ofrecido ningún modelo pasible de alternativa. Esta línea de argumentación inspirada en Monod para la ineludible biologicalidad de los poderes mentales es entonces inconvincente en el mejor de los casos y muy probablemente sea olvidada.

No obstante, es bastante instructiva. Está muy lejos de haberse establecido que los nodos en el sistema masivamente paralelo de alguien *deban* ser neuronas con la materia correcta dentro de ellas, pero sin embargo éste podría ser el caso. Hay otras maneras en las cuales podría probar ser el caso que la reproducción inorgánica de las funciones de procesamiento de información *cognitivamente esenciales* del cerebro humano, tendrían que correr más despacio que sus inspiraciones de tiempo real. Después de todo, hemos descubierto muchos procesos complejos —tales como el tiempo meteorológico— que no puede simularse exactamente en el tiempo real (en el tiempo del pronóstico meteorológico útil, por ejemplo) ni siquiera por las más veloces y más grandes superordenadores que existen actualmente. (No se trata de que las ecuaciones que gobiernan las transiciones no se entiendan. Ni siquiera utilizando lo que ahora sabemos es imposible.) Las ideas geniales pueden muy bien probar ser igualmente difíciles de simular y por tanto predecir. Si lo son, entonces, puesto que la velocidad de la operación es verdaderamente crítica para la inteligencia, tener sólo el programa correcto no es sufi-

ciente, a menos que por "programa correcto" entendamos un programa que pueda correr a la velocidad adecuada para tratar con las entradas y salidas a medida en que se presentan. (Imaginemos a alguien que defendiera la viabilidad del proyecto SDI de Reagan, insistiendo en que el software de control necesario —el "programa correcto"— podría sin ninguna duda escribirse, ¡pero que corriera mil veces más despacio para ser de alguna utilidad en interceptar misiles!)

De manera que si el procesamiento de información en el cerebro realmente se vale totalmente de la velocidad de la actividad computacional de nivel molecular, entonces el precio pagado (en velocidad) por *cualquier* sustitución de material o arquitectura probará ser demasiado elevado. A esta luz, vuélvase a considerar la proposición (D): (D) No hay manera en que se pudiera programar un ordenador electrónico digital de manera que produjera lo que un cerebro humano orgánico con sus poderes causales especiales puede producir de manera demostrable: el control de la actividad intencional veloz e inteligente exhibida por los seres humanos normales.

(D) podría tener razón por el motivo enteramente carente de misterio de que ningún ordenador digital electrónico así podría manejar "el programa adecuado" bastante rápidamente para reproducir el virtuosismo de tiempo real del cerebro. Este es un argumento casi abrumador, pero el punto importante es que sería tonto apostar en contra de él, puesto que podría resultar ser cierto.

¿Pero no es ésta justamente la apuesta que la IA ha hecho? No del todo, aunque algunos entusiastas de la IA estaban sin duda comprometidos con ella. En primer lugar, hasta donde consideremos que la IA es una ciencia, preocupada por desarrollar y confirmar teorías acerca de la naturaleza de la inteligencia o de la mente, la perspectiva de que los ordenadores digitales reales no pudieran correr lo bastante rápidamente como para ser utilizables en nuestro mundo como inteligencias auténticas resultaría de una importancia menor e indirecta, una limitación seria a la habilidad de los investigadores para llevar a cabo experimentos realistas que verificaran sus teorías, pero nada que socavara su afirmación de que habían revelado la esencia de la mentalidad. Hasta donde consideramos a la IA como ingeniería práctica, por otra parte, esta perspectiva sería aplastante para quienes confían tenazmente en crear de verdad una inteligencia humanoide controlada por un ordenador digital, pero esa hazaña es tan irrelevante teóricamente como el malabarismo de construir una vesícula biliar con átomos. Nuestra incapacidad para alcanzar esas metas tecnológicas carece de interés científico o filosófico.

Pero esta manera tan legítima que tiene la IA de apartar con un encogimiento de hombros la perspectiva de que (D) pudiera ser verdad encubre una razón más interesante por la cual la gente de IA podría confiar razonablemente en que mis especulaciones bioquímicas sean una falsificación resonante. Como cualquier esfuerzo en la creación científica, la creación en IA se ha intentado con el carácter de una sobresimplificación oportunista. Es posible acercarse en forma útil y reveladora a cosas que son horriblemente complicadas mediante el uso de separaciones, términos medios, idealizaciones y otras supersimplificaciones deliberadas, con la esperanza de que alguna conducta molar del fenómeno complejo demuestre ser relativamente independiente

de toda la miriada de microdetalles y, por tanto, se reproduzca en un modelo que encubra esos microdetalles. Supongamos, por ejemplo, que un modelo de IA de, digamos, un plan de acción, exija un algún punto, que se consulte un subsistema de visión para obtener información acerca del trazado del ambiente. Más que intentar copiar todo el sistema visual, cuyo funcionamiento es sin duda masivamente paralelo y cuyos productos de salida son evidentemente de una información muy voluminosa, los diseñadores del sistema insertan una especie de doble barato: un "oráculo" clarividente que le puede suministrar al supersistema, digamos, cualquiera de los únicos 256 "informes" diferentes acerca del trazado pertinente del ambiente. Los diseñadores están apostando a que pueden diseñar un sistema para planear actos que se aproximará a la competencia de metas (tal vez la competencia de un niño de cinco años o la de un perro, no la de un adulto maduro) mientras se vale de sólo ocho bits de información visual acerca del trazado ambiental. ¿Es ésta una buena apuesta? Tal vez sí y tal vez no. Hay muchas pruebas de que los seres humanos simplifican sus tareas de manipulación de información y se valen de una fracción diminuta de la información que pueden obtener con sus sentidos. Si esta supersimplificación *especial* resulta ser una mala apuesta, no significará nada más que debemos buscar otra sobresimplificación.

No es en absoluto evidente que alguna combinación unida de las clases de modelos y subsistemas simplificados desarrollados hasta ahora en IA puedan aproximarse a la conducta lúdica de un ser humano normal —ni en el tiempo real ni en órdenes de magnitud menor— sin embargo, eso no pone en tela de juicio la metodología de investigación de IA, más que lo que su incapacidad para predecir el estado del tiempo del mundo real pone en tela de juicio todas las supersimplificaciones meteorológicas como modelos científicos. Si los modelos de IA tienen que modelar "hasta el fin" bajando al nivel neuronal o subneuronal para lograr buenos resultados, éste será un golpe serio para algunas de las aspiraciones tradicionales de IA de adelantarse en la campaña para entender cómo trabaja la mente, pero otras escuelas del IA, como los nuevos conexionistas o los grupos de procesamiento paralelo distribuido, sugieren que se exija ese detalle de nivel bajo para producir una inteligencia artificial práctica significativa en las mentes artificiales. Esta división de opinión *dentro* de IA es radical e importante. Los nuevos conexionistas, por ejemplo, quedan tan claramente fuera de los límites de la escuela tradicional que Haugeland, en *Artificial Intelligence: the Very Idea* (1985), se ve obligado a inventar una sigla, GOFAI (Good Old Fashioned Artificial Intelligence) [La Buena y Anticuada Inteligencia Artificial] para el punto de vista tradicional, en el cual este libro está muy interesado.

¿He pasado ahora tranquilamente en efecto a una defensa de la "IA débil": la simple creación o simulación de los fenómenos psicológicos o mentales por ordenador, como lo contrario de la creación de fenómenos mentales auténticos (pero artificiales) por ordenador? Searle no tiene ningún informe en contra de lo que llama IA débil: "Quizás éste sea un buen lugar donde expresar mi entusiasmo por las perspectivas de la IA débil, del uso del ordenador como herramienta para el estudio de la mente" (Searle, 1982, pág. 57). A lo que se opone es a "la creencia de la IA fuerte de que el ordenador

programado con propiedad tiene literalmente una mente, y a su afirmación antibiológica de que la neurofisiología específica del cerebro no es pertinente al estudio de la mente" (pág. 57). Hay varias maneras de interpretar esta caracterización de la IA fuerte. Creo que la versión siguiente obtendrá la aprobación de la mayor parte de los adeptos:

> La única aplicabilidad de la "neurofisiología específica del cerebro" es la de proporcionar la clase correcta de ingeniería de hardware para la inteligencia del tiempo real. Si resulta que podemos conseguir la velocidad suficiente de las arquitecturas de los microchips de silicio paralelos, la neurofisiología será realmente prescindible, aunque ciertamente valiosa por la sugerencia que puede dar acerca de la arquitectura.
>
> Considérense dos ejecuciones diferentes de un mismo programa, es decir, considérense dos sistemas físicos diferentes, cuyas respectivas transiciones se pueden describir de manera exacta y apropiada en los términos de un único programa "formal" pero uno de los cuales corre seis órdenes de magnitud (alrededor de un millón de veces) más despacio que el otro. (Tomando prestado el ejemplo favorito de Searle, podemos imaginarnos que el lento está hecho con latas de cerveza atadas con un cordel.) En un sentido, ambas ejecuciones tienen las mismas capacidades —ambas "computan la misma función"— pero que en virtud de nada más que su mayor velocidad uno de ellos tendrá "poderes causales" de los que el otro carece: es decir, los poderes de *control* causal que guían un cuerpo locomotor por el mundo real. Por esta misma razón podemos afirmar que el rápido era "literalmente una mente", al mismo tiempo que le negamos ese tratamiento honorífico a su gemelo lento. No se trata de que la mera velocidad (¿velocidad intrínseca?) por encima de cierto nivel crítico cree algún efecto emergente misterioso, sino de que la velocidad relativa es crucial para permitir que las clases correctas de secuencias de interacción ambiente-organismo, se produzcan. Se podría lograr el mismo efecto "aminorando la velocidad del mundo exterior" suficientemente si eso tuviera algún sentido. Un ordenador adecuadamente programado —siempre que sea bastante veloz para intercalarse con los transductores sensoriales y los efectores motores de un "cuerpo" (robótico u orgánico)— tiene literalmente una mente, cualquiera que sea su concreción material, orgánica o inorgánica.

Afirmo que esto es todo con lo que la IA fuerte está comprometida y Searle no ha dado ninguna razón para dudarlo. Vemos como todavía podría ser cierto, como lo proclama la proposición (D), que hay una sola forma de desollar el gato mental después de todo, y es con el verdadero tejido nervioso orgánico. Parecería, entonces, que el tema que separa a Searle de la IA fuerte y sus defensores es una diferencia más bien insignificante de opinión acerca del papel exacto a ser jugado por los detalles de la neurofisiología; pero no es así.

Un par de concesiones hechas a menudo por Searle revelan una diferencia dramática en la implicancia entre las proposiciones (S) y (D). Primero, concede que "casi cualquier sistema tiene un nivel de descripción donde se lo puede describir como un ordenador digital. Se lo puede describir como la concreción de un programa formal. Así que en ese sentido, supongo, todos

nuestros cerebros son ordenadores digitales" (Searle, 1984, pág. 153). Segundo, ha admitido con frecuencia que, por todo lo que él sabe, se podría crear un dispositivo parecido a un cerebro con chips de silicio (u otro hardware aprobado por IA) que imite perfectamente la conducta de entrada y salida del tiempo real del cerebro humano. (Acabamos de dar razones para dudar de lo que Searle concede aquí).

Pero aun si semejante dispositivo tuviera exactamente la misma descripción en el nivel del programa o en el del ordenador digital del cerebro cuya conducta de entrada y salida imitada (en el tiempo real) esto no nos daría *ninguna razón* —según Searle— para suponer que él, como el cerebro orgánico podría de verdad "producir intencionalidad". Si esa imitación perfecta de las funciones de control del cerebro no establecieran que el artefacto era (o "causaba" o "producía") una mente, ¿qué podría arrojar alguna luz sobre este punto, en opinión de Searle? El dice que es una pregunta empírica, pero no dice cómo, ni siquiera en principio, se pondría a investigarla.

Esta es una afirmación enigmática. Aunque muchos (tanto sus críticos como sus seguidores) lo han interpretado mal. Searle insiste en que nunca alegó demostrar que un cerebro orgánico es esencial para la intencionalidad. *Usted* sabe (por algún tipo de relación inmediata, aparentemente) que su cerebro "produce intencionalidad", no importa de lo que esté hecho. Nada en su experiencia directa de intencionalidad podría decirle que su cerebro *no* está hecho de chips de silicio, puesto que "imagine ahora mismo que le abren la cabeza y que adentro no se encuentran neuronas sino otra cosa, digamos chips de silicio. No hay compulsiones puramente lógicas que excluyan algún tipo determinado de sustancia con anticipación" (de próxima aparición, manusc., pág., 1). Searle insiste en que es una pregunta empírica si los chips de silicio producen su intencionalidad, y ese descubrimiento quirúrgico dejaría sentado *para usted* que los chips de silicio podrían de verdad producir intencionalidad, pero no lo dejaría sentado para nadie más. ¿Y si le abriéramos la cabeza a un tercero y le encontráramos chips de silicio? El hecho de que imitaran perfectamente los poderes de control de tiempo real de un cerebro humano no nos daría ninguna razón, en opinión de Searle, para suponer que la tercera persona tenía una mente, puesto que "los poderes de control son irrelevantes por sí mismos" (comunicación personal).

Searle insiste en que es un hecho empírico muy evidente que los cerebros orgánicos pueden producir intencionalidad. Pero uno se pregunta cómo puede haber determinado este hecho empírico. ¡Quizá sólo algunos cerebros orgánicos producen intencionalidad! ¡Tal vez los cerebros de los zurdos, por ejemplo, sólo imitan los poderes de control de los cerebros que producen intencionalidad auténtica! (véase Hofstadter y Dennett, pág. 77). Preguntar a los zurdos si tienen mentes no sirve de ayuda, por supuesto, puesto que sus cerebros pueden no ser más que cuartos chinos.

Con seguridad, es una clase extraña de pregunta empírica que está sistemáticamente privada de toda evidencia empírita intersubjetiva. De manera que la posición de Searle acerca de la importancia de la neurofisiología es que, si bien es importante, por cierto de suma importancia, su contribución crucial podría ser completamente imposible de descubrir desde el afuera. Un cuerpo humano sin una mente verdadera, sin intencionalidad auténtica,

podría bastarse a sí mismo en el mundo real tan bien como un cuerpo humano con una mente real.

Mi posición, por otra parte, como partidario de la proposición (D) es que la neurofisiología es (probablemente) tan importante que si alguna vez veo alguna entidad callejeando por el mundo con la inteligencia de tiempo real de, digamos, C3PO en "La guerra de las galaxias" estaré listo para apostar una suma considerable a que está controlada —en forma local o remota— por un cerebro orgánico. Ninguna otra cosa (lo apuesto) puede controlar una conducta tan inteligente en el tiempo real.

Eso me convierte en una "conductista" a los ojos de Searle y esta clase de conductismo es el meollo del desacuerdo entre la IA y Searle. Pero éste es el "conductismo" blando de las ciencias físicas en general. No se trata de ningún dogma skinneriano o watsoniano (o ryleano) limitado. La conducta en este sentido blando incluye todos los procesos y hechos internos observables intersubjetivamente (como la conducta de sus entrañas o su ARN). Nadie se queja de que los modelos de la ciencia sólo expliquen la "conducta" de los huracanes o las vesículas biliares o los sistemas solares. ¿Qué más hay acerca de estos fenómenos que la ciencia tenga que explicar? Esto es lo que hace tan misteriosos a los poderes causales que Searle imagina: no tienen, según él mismo lo admite, ningún efecto delator sobre la conducta (interna o externa) —a diferencia de los poderes causales que yo me tomo tan en serio: los poderes necesarios para guiar a un cuerpo por la vida, ver, oír, moverse, hablar, decidir, investigar, etcétera. Es, por lo menos, engañoso llamar a una doctrina tan completamente cognitiva y (por ejemplo) antiskinneriana como la mía conductismo, pero Searle insiste en usar el término de este modo.

Repasemos la situación, Searle critica a la IA por no tomar en serio a la neurofisiología y la bioquímica. He sugerido una manera en la cual la bioquímica del cerebro podría, ciertamente, jugar un papel crítico: suministrando la velocidad de funcionamiento para el pensamiento veloz, pero ésta no es la clase de poder bioquímico causal en el que Searle piensa. Supone que hay una "distinción clara entre los poderes causales del cerebro para producir estados mentales y los poderes causales del cerebro (junto con el resto del sistema nervioso) para producir relaciones de entrada y salida" (de próxima aparición, manusc., pág. 4). Llama a los primeros "los poderes causales de abajo a arriba del cerebro", y sólo lleva "un momento de reflexión" ver la falsedad en la idea de que los segundos so los que importan para la mentalidad: "la presencia de la causación de entrada y salida que le permitiría a un robot funcionar en el mundo *no implica nada en absoluto* acerca de la causación de abajo a arriba que produciría estados mentales". El robot triunfador "podría ser un zombi total" (de próxima aparición, manusc., pág. 5).

De manera que ése es el punto crucial para Searle: la conciencia, no "la semántica". Su punto de vista no se apoya, como él dice, en "el concepto moderno de computación y por cierto en nuestra visión moderna del mundo científico" sino en la idea, que él cree confirmable por cualquiera que tenga un momento libre para reflexionar, de que la IA fuerte no lograría distinguir entre un "zombi total" y un ser con intencionalidad real, intrínseca. La conciencia introspectiva, *cómo es* ser usted (y entender el chino) es el verdadero tema de Searle. A pesar de su insistencia en que es muy simple argumento es el centro de mesa de su punto de vista y que la "parábola" del cuarto chino es

sólo un recordatorio vívido de la verdad de su segunda premisa, su caso depende en realidad del "punto de vista de la primera persona" del hombre que está en el cuarto chino.

Aparentemente, Searle ha confundido una afirmación acerca de la inderivabilidad de *la semántica* de la sintaxis con una afirmación acerca de la inderivabilidad de la *conciencia de la semántica* de la síntaxis. Para Searle, la idea de la comprensión auténtica, la "semanticidad" auténtica como a menudo la llama, es inexplicable de la idea de la conciencia. Ni siquiera considera la posibilidad de la semanticidad inconsciente.

Los problemas de la conciencia son serios y desconcertantes para la IA y para todos los demás. La cuestión de si una máquina *podría ser consciente* ya la he tratado en detalle antes (*Brainstorms*, b capítulos 8-11; Hofstadter y Dennett, 1981; Dennett, 1982b, 1985a, de próxima aparición e) y me ocuparé de ella con detalle en el futuro. Este no es ni el tiempo ni el lugar para una discusión en gran escala. Por el momento, limitémonos a notar que el planteamientos de Searle, tal como está no depende en absoluto del muy simple argumento acerca de la formalidad de los programas ni de la derivabilidad de la semántica de la sintaxis sino de intuiciones profundamente arraigadas que la mayoría de la gente tiene de la conciencia y su aparente imposibilidad de ser factible en una máquina.

El tratamiento que Searle le da a ese caso nos invita, además, a retroceder a una posición cartesiana ventajosa. (La furia de Searle no es nunca más feroz que cuando un crítico lo llama dualista, porque él sostiene que es un materialista completamente moderno; pero entre sus principales partidarios, que creen estar de acuerco con él, hay cartesianos actuales como Eccles y Puccetti.) Searle proclama que de algún modo —y no tiene nada que decir acerca de los detalles— la bioquímica del cerebro humano asegura que ningún ser humano sea un zombi. Esto es tranquilizador pero engañoso. ¿Cómo crea la bioquímica un efecto tan feliz? Por cierto que mediante un poder causal admirable. Es el mismo poder causal que Descartes les imputaba a las almas inmateriales y Searle no lo ha hecho menos maravilloso o misterioso —o incoherente al fin— asegurándonos que de algún modo sólo es un problema de bioquímica.

Finalmente, para responder abiertamente al desafío de Searle: ¿qué pienso que está mal en el argumento tan simple de Searle aparte de ser una pista falsa? Tomemos en cuenta una vez más su

> Proposición 2: La sintaxis no es equivalente para la semántica, ni suficiente por sí misma.

Se puede afirmar que esto todavía es cierto si cometemos el sencillo error de hablar de la sintaxis en un estante, en un programa no puesto en práctica. Pero la síntaxis *encarnada, en actividad* —el "programa correcto" en una máquina adecuadamente veloz— *es* suficiente para la intencionalidad *derivada*, y ésa es la única clase de semántica que hay, como lo he explicado en el capítulo 8 (véase también la discusión de los mecanismos sintácticos y semánticos en el capítulo 3). De manera que rechazo, con argumentos, la proposición 2 de Searle.

En realidad, las mismas consideraciones muestran que hay también algo fuera de lugar en su proposición 1: los programas *son* puramente formales (es decir, sintácticos).

La cuestión de si un programa ha de ser identificado por sus características puramente formales es debatido muy acaloradamente en estos días por la ley. ¿Se puede patentar un programa o siquiera registrarlo como propiedad intelectual? Un sinnúmero de pleitos interesantes pululan alrededor de la pregunta de si los programas que ***hacen las mismas cosas en el mundo*** cuentan como el mismo programa aun cuando sean en cierto nivel de descripción, sintácticamente diferentes. Si los detalles de la "encarnación" se incluyen en la especificación de un programa y se consideran esenciales para él, entonces el programa no es en absoluto un objeto puramente formal (y es perfectamente elegible para ser protegido por una patente), y si no se fijan ***ciertos*** detalles de la encarnación —por la semántica interna del lenguaje de la máquina en el cual el programa está finalmente escrito— un programa no es siquiera un objeto sintáctico sino sólo un modelo de marcas tan inerte como un papel de empapelar.

Finalmente, una implicación de los argumentos del capítulo 8 es que la proposición 3 de Searle es falsa, dado lo que él quiere decir con "las mentes tienen contenidos mentales". No existe algo como la intencionalidad intrínseca, en especial si se la considera como Searle ahora lo requiere, como una propiedad a la que el sujeto tiene acceso consciente y privilegiado.

10

Examen de mitad de curso: Comparación y contraste

P.: *Compare y contraste las ideas de los siguientes filósofos sobre el status fundamental de las atribuciones de intencionalidad: Quine, Sellars, Chisholm, Putnam, Davidson, Dennett, Fodor, Stich y Dennett.*

R: Sería extraño y angustioso que las "diferencias más importantes" entre estos teóricos filosóficos resultaran ser tan importantes —si, en particular, resultara que un aspecto de cada controversia estuviera totalmente equivocado sobre algo importante (en contraste con, digamos, haber puesto demasiado énfasis en algún aspecto de la verdad). Sería angustioso porque simplemente no sería cuestión de que un grupo de gente tan inteligente pudiera leer y analizar los mismos libros, trabajara en la misma tradición (angloamericana), estuviera familiarizada con aproximadamente la misma evidencia, respaldara la misma metodología y, no obstante, algunos de ellos no lograran entender totalmente el significado de todo esto, a pesar de los esfuerzos más sabios de sus colegas por esclarecérselos. Todos cometemos errores, por supuesto, y nadie es inmune a la confusión, pero a menos que la filosofía sólo sea un juego de incautos, como opinan algunos de sus detractores, y si todas las otras cuestiones son iguales, sería de esperar ver que estos teóricos hicieran causa común, contribuyendo cada uno de ellos con algo para la obtención de un esclarecimiento común emergente sobre la naturaleza de la mente y su relación con el cuerpo y el resto del mundo físico.

Felizmente, en este caso existe una perspectiva desde la cual predomina el acuerdo, se puede discernir el progreso, y muchas de las oposiciones más salientes parecen ser los productos amplificados de diferencias menores de juicio o gusto, o de lo que podría llamarse aseveración exagerada táctica. (Si creemos que nuestro interlocutor se está perdiendo algo importante, tendemos a exagerar su importancia en nuestro esfuerzo por remediar el equilibrio. Además, la mejor forma de lograr que otros nos presten atención es decir algo memorablemente "radical", con la menor cantidad posible de calificativos y concesiones hipócritas.) A nadie le agrada enterarse de que su carrera de cruzado no es más que una parte de tormenta en un vaso de agua, y así es típico que los filósofos no se sientan ansiosos por aceptar resoluciones tan insulsas y ecuménicas de sus polémicas, pero puede resultar tranquilizador y hasta esclarecedor —aunque no excitante en especial— para nosotros mismos, tener presente qué grado exacto de acuerdo fundamental existe.

Nos ayudará a comprender el estado actual sobre esta cuestión volver atrás unos pocos años a los textos que fija el orden del día: Chisholm, en *"Sentences about Believing"* (1956) (véase también Chisholm, 1957), observó tanto el estrépito lógico como la irreductibilidad aparente de los modismos intencionales y pasó a plantear el dilema que estos modismos formulaban para aquellos que deseaban unificar la mente y la ciencia. Chisholm observó que algunos filósofos y psicólogos

> parecen haber considerado que sería filosóficamente significativo que pudieran demostrar que las oraciones de creencia pueden ser reescritas en un idioma adecuado que no es intencional o que, por lo menos, sería significativo demostrar que [la tesis de irreductibilidad de] Brentano estaba mal. Supongamos por un momento que *no podemos* reescribir oraciones de creencia en forma tal que sea contraria a nuestra versión lingüística de la tesis de Brentano. ¿Cuál sería el significado de este hecho? Considero que esta pregunta es en sí misma filosóficamente significativa, pero no estoy preparado para responderla (pág. 519).

Quine, trabajando desde un punto de vista notablemente diferente (su celebrada exploración de la tarea de "traducción radical"), llegó al mismo veredicto: no hay forma de reducir estrictamente o de traducir los modismos de significado (o la semántica o la intencionalidad) al idioma de la ciencia física. Observó la convergencia. "La tesis de la irreductibilidad de los modismos intencionales de Brentano es consistente con la tesis de la indeterminación de la traducción" (1960, pág. 221). No es sorprendente que también haya estado de acuerdo con Chisholm sobre la importancia filosófica de la pregunta que Chisholm no estaba preparado para responder. Pero la unificación de la ciencia siempre ha cobrado suma importancia como meta para Quine (aun cuando no haya parecido conmover especialmente a Chisholm), de modo que él estaba en verdad preparado para responder a esa pregunta. En un pasaje citado muy a menudo, Quine declaró valientemente su fidelidad a una clase estricta de conductismo (un ejemplo de la doctrina llamada más recientemente *materialismo eliminativo):*

> Se puede aceptar la tesis de Brentano, sea como muestra de la indispensabilidad de los modismos intencionales y la importancia de una ciencia autónoma de intención, o como indicación de la falta de fundamento de los modismos intencionales o la vacuidad de una ciencia de la intención. Mi actitud, a diferencia de la de Brentano, es la segunda (1960, pág. 221).

De hecho, estos dos pasajes son ligeramente engañosos, en esto que Chisholm, a pesar de su anunciada reticencia, ha actuado en forma heroica durante años en defensa del polo mentalista de Brentano (llamémoslo Norte)[1] y Quine, a pesar de su voto solemne y sonoro de conductismo (Sur) polar, desde el comienzo ha estado explorando y defendiendo un territorio algo más moderado, junto con todos los demás. Anscombe (1957), Geach

[1] No confundir con el Polo Este, que fija una geografía lógica diferente en Dennett, 1984b, de próxima aparición e.

(1957) y Taylor (1964) efectuaron exploraciones nórdicas notables que fueron recibidas con beneplácito y alternativas de importancia al rechazo de Quine de lo mental y pronto obtuvo aceptación común entre los filósofos el hecho de que el fenómeno del significado o contenido de los estados psicológicos —intencionalidad— no era tan inasequible a las ciencias físicas como Brentano y Chisholm parecían sugerir, ni tan confortablemente desechable como Quine parecía sugerir. Debe ser posible alguna clase de avenencia unificadora entre una "reducción" rigurosa de lo mentalista o lo brutalmente fisicalista y una negación abrupta de los fenómenos de la mente.

El Ecuador, para continuar con la metáfora geográfica, ya había sido delimitado por Sellars, en su importante correspondencia con Chisholm:

> Mi solución es que "'...'significa---" es el núcleo de un modo único de discurso que es tan diferentes de la *descripción* y *explicación* del hecho empírico, como lo es el idioma de la *prescripción* y la *justificación* (Chisholm y Sellars, 1958, pág. 527).

Lo que era singular en este modo de discurso, según los análisis precursores de Sellars (1954, 1956, 1963), era su apelación ineliminable a las consideraciones funcionales. Así nació el *funcionalismo* contemporáneo en la filosofía de la mente, y las variedades de funcionalismo que hemos visto con posterioridad son habilitadas de una u otra forma, e inspiradas directa o indirectamente, por lo que dejó abierto la propuesta inicial de Sellars— aunque esto no ha sido ampliamente aceptado. [2]

Pero Quine, tal como se acaba de hacer notar, no estaba en realidad ocupando el árido Polo Sur de su notorio conductismo, a pesar de lo que decía a menudo.

El, junto con Sellars, aunque levemente al sur de éste, estaba fijando algunos de los límites de toda componenda funcionalista. La sección de *Word and Object* adecuadamente titulada "The Double Standard" (1960, páginas 216-21), puede leerse hoy como una valiosa presentación previa de las contribuciones de Putnam, Davidson, Bennett, Stich y Dennett, entre otros.

El doble modelo apoyado por Quine dependía de la *seriedad* con que se suponía que se tomaban los modismos intencionales.

> Si estamos describiendo la estructura verdadera y fundamental de la realidad, el esquema canónico para nosotros es el esquema austero que no conoce otra cita más que la cita directa y no conoce ninguna actitud proposicional, sino únicamente la constitución física y la conducta de los organismos (pág. 221).

En otras palabras, hablando estrictamente, hablando ontológicamente, no existen cosas como las creencias, los deseos u otros fenómenos intenciona-

[2] La influencia de Sellars ha sido ubicua, pero casi subliminal (si hay que juzgar por la escasez de citas a Sellars entre los funcionalistas). Es evidente que Putnam, Harman y Lycan (1974, 1981a, 1981b) han sido influidos muy directamente por Sellars, pero Dennett, Fodor, Block y Lewis demuestran que la influencia de Sellars es, en gran parte de segunda mano, y principalmente por la vía de la serie de artículos muy influyentes de Putnam, reimpresos en Putnam, 1975b. El papel de Sellars en la evolución del funcionalismo se aclara en los comentarios de Putnam (1974) y los de Dennett (1974a) sobre "Meaning as Functional Classification (A Perspective on the Rela-

les. Pero los modismos intencionales son "prácticamente indispensables" y deberíamos ver lo que podemos hacer para encontrar algún sentido a su empleo en lo que Quine llamó un modismo "esencialmente dramático" (pág. 219). Luego, en todo uso del vocabulario intencional, deben reconocerse no simplemente los hechos en bruto, sino un elemento de interpretación, y más aun, de interpretación dramática.

Aquí encontramos a Quine y Sellars en un acuerdo fundamental acerca de la naturaleza no simplemente descriptiva de la atribución intencional, y casi todos han coincidido, desde entonces, aunque con distinto énfasis. Por ejemplo, la mayoría ha considerado que la afirmación de Quine de que estos modismos son simplemente *"casi"* indispensables, subestimó la centralidad del papel que desempeñan, pero si suprimimos el "simplemente" podría quedar poco para debatir.

Lo que ha quedado como asunto de arduo debate desde entonces es cómo jugar este juego de interpretación dramática. ¿Cuáles son los principios de la interpretación y sus presuposiciones e implicancias? Aquí parecen haber emergido dos rivales principales: uno u otro principio normativo, según el cual se deben atribuir a un ser las actitudes proposicionales que "debería tener" dadas sus circunstancias, y uno u otro Principio Proyectivo, según el cual se le deben atribuir a un ser las actitudes proposicionales que se supone que uno mismo debería tener en esas circunstancias.

Con el principio normativo hemos tenido las diversas subvariedades del principio de caridad (Davidson, 1967, 1973, 1974, recopiladas con otras en 1985; y Lewis, 1974) y la presunción de racionalidad (Dennett, 1969, 1971, 1975; Cherniak, 1981, 1986). Mucho de lo que Davidson y Dennett entendieron de este tema estaba preanunciado en *Rationality* y de Bennet (1964) —aunque se necesita cierta visión retrospectiva para apreciar el alcance de esto— pero en cualquier caso, todo se origina en el análisis de Quine de la necesidad de un principio semejante en cualquier ejercicio de traducción radical.

> La traducción caprichosa impone su lógica sobre ellos, y rogaría por la pregunta de prelogicalidad si hubiera una pregunta por la cual rogar... El axioma de la traducción que subyace a todo esto es la probabilidad de que las afirmaciones sorprendentemente falsas a primera vista resulten ser diferencias ocultas del idioma... El sentido común que respalda el axioma es que la tontería de nuestro interlocutor, más allá de cierto punto, es menos probable que una mala traducción o, en el caso doméstico, que la divergencia lingüística (1960, pág. 59).

(En una nota al pie de página, Quine atribuye a Wilson, 1959, el mérito de la idea de un principio de caridad, con ese nombre.) De modo que Quine

---

tion of Syntax to Semanties)" de Sellars, y en la respuesta de éste, en la conferencia sobre intencionalidad, lenguaje y traducción dictada en la Universidad de Connecticut en marzo de 1973. El curso de acción de esta conferencia, publicado como edición especial de *Synthese* (1974) también constituye un tesoro de discernimientos sobre la tesis de Quine de la indeterminación de la traducción radical y su relación con el problema de la intencionalidad en la filosofía de la mente. Véase también Harman, 1968, 1986 y Lucan, 1981b para una mayor aclaración de esta historia.

es el padre del principio normativo, salvo que pueda hallarse una apreciación contemporánea del papel de tales consideraciones normativas en los papeles funcionales individualizadores en Sellers.

¿Y qué del principio proyectivo? Grandy (1973) desarrolló una versión inicial, contrastando lo que llamaba el principio de humanidad con el principio de caridad, y Stich (1980, 1981, 1983, 1984) ha ofrecido la defensa más detallada y vigorosa de esta idea, pero Quine también es el padre de este principio (como observa Stich, 1983, pág. 84):

> Cuando citamos directamente las declaraciones de un hombre, las relatamos como lo haríamos con el llamado de un ave. Por significativa que sea la declaración, la cita directa simplemente relata el incidente físico y nos deja todas las implicaciones a nosotros. Por otra parte, en la cita indirecta nos proyectamos dentro de lo que, según sus observaciones y otras indicaciones, nos imaginamos que ha sido el estado de ánimo del que habla, y luego decimos lo que en nuestro idioma es natural y pertinente para nosotros en el estado así simulado (Quine, 1960, pág. 219).

Para complicar aun más las cosas para los taxonomistas, Quine reconoce el mérito de las ideas de su perspicaz análisis de las expectativas de un principio proyectivo, a sus conversaciones con Davidson (Quine, 1960, pág. 217n). Además, el análisis elaborado de Sellars (1954) de la clasificación funcional de términos relacionados con las funciones de los términos de ***nuestro*** idioma, debe considerarse como el ***locus classicus*** de la interpretación proyectivista.

Levin (de próxima aparición) divide a los contendientes en esta indagación sobre los principios de la interpretación, en racionalizadores y proyectores (véase también Stich, 1984). Observa la tendencia de las versiones suaves de cada uno a fundirse en una opinión única, pero no lleva este ecumenismo tan lejos como Dennett (capítulo 4 de este libro), que argumenta que la oposición entre proyección y racionalización es a lo sumo una cuestión de énfasis. Como Quine observó desde el comienzo, el problema con la estrategia de la proyección es que "Al poner así a nuestras identidades reales a interpretar papeles irreales, por lo general no sabemos cuánta realidad mantener constantemente. Surgen así dilemas" (1960, pág. 219). Exactamente los dilemas que Stich (1983) examina en detalle; dilemas que son resolubles —hasta el punto en que lo son— sólo recurriendo a consideraciones normativas: deberíamos proyectar sólo lo mejor de nosotros mismos, pero lo que se estima mejor bajo las circunstancias es en sí mismo cuestión de interpretación.

Ahora bien, si se tomara en serio —muy en serio— la atribución de actitudes proposicionales por la vía de los modismos intencionales, habría ahí un problema real: exactamente qué principio o cuáles principios de interpretación dan las actitudes proposicionales verdaderas o ***reales***. Pero para Quine, como hemos visto, este problema no surge, ya que el "modismo dramático" no es más que una necesidad práctica de la vida diaria, sujeta a consideraciones puramente pragmáticas, y no una forma de delinear la realidad fundamental.

> Por lo común, el grado de desviación permisible depende de por qué citamos un texto. Se trata de qué rasgos de las observaciones del autor citado deseamos descifrar; ésos son los rasgos que debemos tener en claro si deseamos que nuestra cita indirecta cuente como cierta. Para las oraciones de creencia y otras actitudes proposicionales son válidas observaciones similares... También sucede a menudo que simplemente no hay forma de decir que una afirmación de la actitud proposicional es verdadera o falsa, aun dado un conocimiento pleno de sus circunstancias y propósitos (1960, pág. 218).

Según Quine explicó en 1970:

> La metáfora de la caja negra, a menudo tan útil, puede aquí conducir a error. El problema trata de hechos ocultos, como podrían descubrirse aprendiendo algo más acerca de la filosofía cerebral de los procesos del pensamiento. Esperar un mecanismo físico distintivo detrás de todo estado mental genuinamente preciso es una cosa; esperar un mecanismo distintivo para toda distinción expresable que pueda formularse en idioma mentalista tradicional, es otra. La cuestión de si... el extranjero cree *en realidad* (a *A* o cree más bien a *B*, es una cuestión cuya significación misma pondría en duda. Esto es lo que estoy queriendo dar a entender al argumentar sobre la indeterminación de la traducción (páginas 180-81).

Pasemos revista a cómo llegamos tan lejos. Casi todos aceptan la tesis de irreductibilidad de Brentano, pero si se la acepta fundamentalmente por la razones de Quine —porque se ha visto que existe indeterminación en la traducción radical— no se estará propenso a ser un realista (estricto) acerca de las atribuciones de la actitud proposicional, y por tanto no tendrán tendencia a ser realista acerca del contenido psicológico (intencionalidad intrínseca o genuina). Como dice Quine: "Aceptar el uso intencional por su valor nominal es, según hemos visto, postular las relaciones de la traducción como algo objetivamente válido, aunque indeterminado en principio, en relación con la totalidad de las disposiciones del idioma" (1960, pág. 221). Gran parte del debate sobre los principios de la interpretación ha sido desplazado y deformado porque los participantes no han podido ponerse de acuerdo sobre esta implicación de la opinión de Quine. (El único análisis explícito de estas implicancias es el que encontré en Lycan, 1981b). Así, para Davidson y Dennett, que son los quineanos más entusiastas en este aspecto, simplemente no existe tema de discusión alguno, allí mismo donde aparece amenazante una gran laguna para los realistas como Fodor (y Burge y Dretske y Kripke y otros —véase el capítulo 8—). Las oscilaciones a observar en las opiniones de Putnam a lo largo de los años [1974, 1975b, 1978 (ver en especial su Conferencia IV sobre John Locke, páginas 54-60), 1981, 1983, 1986] descubren sus exploraciones (Norte y Sur) de los costes y beneficios de concordar con Quine sobre la indeterminación. En sus trabajos más recientes, concluye más o menos con Quine y los quineanos:

> La explicación de creencia/deseo pertenece al nivel de lo que he estado llamando *teoría de la interpretación*. Es tan holística y de interés relativo, como toda interpretación. Los psicólogos hablan a menudo como si hubiera *conceptos* en

el *cerebro*. El propósito de mi argumento (y, creo, del de Davidson), es que en el cerebro puede haber *análogos de oración* y *análogos de predicado*, pero no conceptos. "Las representaciones mentales" requieren interpretación, exactamente igual que cualquier otro signo (1983, pág. 154).

Mientras tanto, Fodor, alumno de Putman, ha estado siguiendo un rumbo notablemente independiente hacia el norte, defendiendo tanto la irreductibilidad como la realidad de los estados intencionales, pero —a diferencia de Chisholm, Anscombe, Geach y Taylor, por ejemplo— en un intento por volver estas realidades irreductibles aceptables para las ciencias físicas, al fundamentarlas (de alguna manera) en la "sintaxis" de un sistema de representaciones mentales comprendidas físicamente. En *Psychological Explanation* (1968a), uno de los textos definitorios del funcionalismo. Fodor dio la espalda resueltamente a Quine (en sus oropeles conductistas) y a Ryle (en sus oropeles conductistas totalmente diferentes y a Wittgenstein en sus oropeles conductistas aun más diferentes), y comenzó a bosquejar una enumeración de acontecimientos mentales como *procesos internos*, distinguidos o identificados por sus propiedades funcionales. Parecía ser, y se suponía que debía parecerlo, una alternativa totalmente extrema. Parecía ser una forma de socavar lo que en ese momento se manifestaba como un dogma conductista asfixiante: *lo que va por dentro no resuelve nada*.

Mientras *The Language of Thought* (1975) de Fodor desarrolló más aun el tema anti-Ryle y anti-Quine, la teoría positiva no parecía ser, a pesar de todo, o *tan* diferente de la de Sellars, con su reconocimiento de las presuposiciones teleológicas de las taxonomías funcionales. [3] Con anterioridad Sellars había desarrollado su propia alternativa hasta llegar por lo menos a alguna especie de conductismo, en su "Myth of our Rylean Ancestors", 1956, reimpreso en Sellars, 1963). Imaginaba a antepasados con un idioma ryleano (conducista) que llegaban a postular, como entidades teóricas, ciertos episodios internos del lenguaje mentalista, estados que eran identificados por sus papeles funcionales, o —lo que al final resultaba ser la misma cosa— por su significado o intencionalidad (véase también Sellars, 1954).

Tal vez la diferencia crucial entre los funcionalismos del lenguaje mentalista de Sellars y de Fodor reside en que Sellars, en su análisis de "reglas de entrada del idioma" y "reglas de salida del idioma" para el lenguaje mentalista, admitía la necesidad de una *especie* de análisis "conductista" de las propiedades semánticas de esas representaciones interiores —lo que más adelante se denominaría una *semántica del procedimiento*—. Pero a pesar de la primitiva declaración de Fodor de la dependencia (de alguna manera) de la semántica de la función, él ha resistido este paso —dado más explícitamente por Dennett (1969, capítulo 4)— porque amenazaba el papel que esperaba que fuera llenado por el lenguaje mentalista postulado. [4].

[3] Resulta instructivo comparar el libro de Fodor de 1975 con *Throught* (1973) de Harman, una versión del idioma del pensamiento con consonancias marcadas con Quine y Sellars aunque Harman jamás cita a Sellars.

[4] Véase Fodor, 1981e, para una diatriba contra la semántica del procedimiento en nuestros días. Fodor (1975) no cita a Sellars y en 1981a hay una referencia a éste. "Descubro, muy tardíamente, que

Chisholm y Sellars (1958, páginas 544ss.), que en otros aspectos diferían, tenían ambos muy clara la imposibilidad de *fundamentar* el significado en las propiedades semánticas de alguna manera "primitivas" de un idioma "interior". A Fodor, por otra parte, un lenguaje interior del pensamiento le ha seguido pareciendo una forma alternativa y más poderosa de *resolver* cuestiones de interpretación psicológica, más que, como Quine, Sellars, Davidson y Dennett (1973) han insistido siempre (y ahora Putnam está de acuerdo), un simple replanteo del problema de la traducción radical. Fodor ha tenido que virar peligrosamente cerca del polo norte de la intencionalidad real e intrínseca, aun cuando se mofe de ello:

> Mi punto de vista, entonces, no es, desde luego, que el solipsismo es verdadero; es precisamente que la verdad, la referencia y el resto de las ideas semánticas no son categorías psicológicas. Son modos de *Dasein*. No sé lo que es *Dasein*, pero estoy seguro de que haya mucho de eso alrededor y estoy seguro de que usted y yo y Cincinnati, todos lo tenemos. ¿Qué más quieren? (1908, pág. 71; vuelto a imprimir en 1981a).

"Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum" (1985) es la propia respuesta de Fodor a la pregunta de nuestro examen, y presenta una taxonomía entretenida y a menudo perspicaz, si bien algo rígida, como el mismo Fodor admite. Fodor defiende francamente el realismo y presenta el instrumentalismo de Dennett como su "primera opción antirrealista" (pág. 79). No se toma el trabajo de refutar este instrumentalismo, pero luego de haberle asestado algunos mandobles incidentales, lo deja de lado: "Entre una cosa y otra, parece de verdad posible dudar de que esté próximo un instrumentalismo coherente acerca de las actitudes". Luego procede a presentar su taxonomía de los otros caminos, pero lo que impresiona a este lector es que todas las posiciones descendentes de la bifurcación que se aleja de la posición de Dennett —las posiciones a las cuales Fodor dedica el resto de su revisión— están, según su propio relato tan plagadas de dilemas que casi podría suponerse que estaba catalogando un *reductio ad absurdum* involuntario. En particular, Fodor cita lo que denomina el "problema de idealización" —el principio de caridad enmascarado— como un problema no resuelto, y admite que no "ve razón alguna para suponer que el problema pueda resolverse" dentro de sus límites realistas (pág. 97). Concluye con la siguiente observación: "Pero, como están las cosas, no tenemos explicación alguna de la semanticidad de las representaciones mentales" (pág. 99).

Bennett y Stich han explorado otras dos regiones nórdicas, que resisten en sus formas diferentes el clima seductor del Ecuador. Bennett (1976) rechazó explícitamente la tesis de indeterminación de Quine (páginas 257-64) y recurriendo al análisis de explicación teleológica de Taylor (1964) intentó una teoría realista detallada del contenido de lo que llamó *registros*, los estados de representación interior de los animales, de los cuales las *creencias* son las subvariedad

---

Sellars (1956) propuso una vez una explicación en algunos aspectos parecida a ésta. El trabajo de Sellars parece ser notablemente "presciente a la luz de (lo que yo considero que son) las presuposiciones metodológicas de la psicología cognitiva contemporánea" (páginas 325-26).

humana especial. Como Dennett, y a diferencia de Fodor y Davidson, eligió así la estrategia de intentar aclarar primero las necesidades de representación y los recursos de los animales que no utilizan un lenguaje, antes de tratar de elaborar una relación de la creencia humana (o, como Dennett diría, la opinión) sobre esa base. En realidad, los caminos paralelos de las investigaciones de Dennett y Bennett están separados por poco más de su desacuerdo sobre la tesis de la indeterminación y sus implicaciones para las perspectivas de una "teoría subyacente firme" de "estructuras conceptuales" [véase Bennett, (1983), y la respuesta de Dennett (páginas 382-83) en *BBS*].

Podemos ver ahora que Stich (1983), descubrió una curva de vuelta al Ecuador: una ***reductio ad absurdum*** exhaustiva de la hipótesis de que alguna forma del principio estricto de la interpretación proyectivista nos pueda permitir, después de todo, ser realistas acerca del contenido, a pesar de los escrúpulos de conciencia de Quine. Stich termina uniéndose a Quine en su doble modelo: hablando estrictamente, no existen cosas tales como las creencias, aun cuando sea una necesidad práctica el ***hablar como si las hubiera***.

Churchland (1981), que ofreció razones algo distintas, llegó al mismo destino, aunque en el proceso aparentemente disentía sinceramente de Dennett y Putnam. Es éste otro caso de diferencia en el énfasis que aparece a lo lejos como un desacuerdo importante: Churchland, a diferencia de Dennett y de Putnam, no encuentra atractiva la idea quineana del doble modelo. Es decir, no toma en serio la idea de no tomar en serio el tema de conversación de la actitud proposicional, aun cuando sí lo hace. Admite, como todo materialista eliminador sensato que, a los fines prácticos, seguiremos hablando como si hubieran creencias y deseos, metiéndonos en "el modismo dramático", como dice Quine. Pero para Churchland, se trata de una admisión descartable, no del preludio a la teoría de la interpretación que es para Davidson, Dennett y Putnam.

¿Qué otras diferencias quedan entre los teóricos de la interpretación? Dennett y Davidson también han seguido cursos notablemente paralelos pero independientes a través de los años, unidos por su aceptación de la indeterminación quineana y su fidelidad a los principios de caridad de la interpretación. ¿Qué los sigue separando? Haugeland observa:

> Dennett... difiere de Davidson en su manera ontológica de ver las cosas —es decir, en su actitud hacia las identidades consignadas en las estructuras respectivas. Davidson restringe su análisis sólo a acontecimientos, mientras que Dennett también incluye afortunadamente estados, procesos, estructuras y cosas por el estilo. Pero esta discrepancia superficial refleja una divergencia mucho más profunda y más importante. El propósito de Davidson es demostrar que cada acontecimiento mental es exactamente el mismo acontecimiento que algún acontecimiento físico, y su argumento depende de una doctrina de relaciones causales que él aplica sólo a los acontecimientos. Dennett, por otra parte, no sólo no comparte ese fin, sino que aparentemente tampoco aceptaría la conclusión. [Para Dennett], las creencias son especificables... en el mejor de los casos, por una especie de "equilibrio" racional en la estructura intencional; y de ahí que su status como entidades, de ser características, deberían reflejar esta diferencia de estructura. En otras palabras, tal vez Dennett debería estar de acuerdo con Ryle:
>
> Es perfectamente correcto decir, en un tono lógico de voz, que existen mentes, y decir, en otro tono lógico de voz, que existen cuerpos. Pero estas expresio-

nes no indican dos especies diferentes de existencia, porque "existencia" no es una palabra genérica como "de color" o "sexuado" (1949, pág. 23).

Davidson, en efecto, encontraría totalmente inaceptable esta sugerencia. Heidegger, por otra parte, la hallaría totalmente compatible; porque su punto de vista acerca de la presencia-a-la-mano, la disposición-a-la-mano y la existencia es precisamente que son distintas "formas de ser" (Haugeland, inédito, páginas 5-6). [5]

Haugeland tiene razón: Dennett debería estar —y estuvo— de acuerdo con Ryle en este punto exacto (Dennett, 1969, páginas 6-18), al tiempo que intentaba avenirlo con su acuerdo con Quine, sobre las implicaciones de la traducción esencial. Davidson, por otra parte, revela por su realismo *ontológico* firme sobre las creencias (como items que había que esperar incluir en la ontología de la ciencia unificada) que siempre ha querido tomar el doble modelo de Quine con cierto escepticismo. Continúa con su anhelo de tomar las actitudes proposicionales con más seriedad de lo que recomendaría Quine, a pesar de concordar con éste acerca de la indeterminación de la traducción radical. La "divergencia profunda" sobre la visión ontológica que Haugeland discierne correctamente entre Davidson y Dennett, también puede considerarse así simplemente como los efectos ampliados de una diferencia menor de opinión acerca de *exactamente* con cuánta seriedad tomar el doble modelo.

Toda esta convergencia, entre filósofos con actitudes, aspiraciones y métodos totalmente diferentes, robustece la convicción —tanto más aun en vista de los curiosos modelos de *no*— cita en la literatura en estudio. Como ya se hizo notar, Dennett y Davidson, Dennett y Bennett, casi nunca se citan o discuten entre sí, a pesar de que transitan sendas sumamente paralelas a través de territorios aledaños. Casi nadie cita a Sellars, si bien reinventan sus ruedas con regularidad gratificante. Podrían citarse muchos otros ejemplos de reinversión no reconocida. ¿Qué es lo que explica esto? Sospecho que el atraso del tiempo de comprensión. Los filósofos no están jamás totalmente seguros de lo que están hablando —acerca de cuáles son *realmente* los problemas— y así a menudo les lleva un tiempo relativamente largo admitir que alguien con un enfoque (o destino o punto de partida) *algo* diferente, esté efectuando una contribución. Admitimos —y citamos y analizamos— rumbos de colisión frontal con mucha más facilidad que las trayectorias casi paralelas, que tienden a golpearnos, si es que damos cuenta de ellas, como demasiado evidentes para ser comentadas. Además, no es como si los filósofos descubrieran sus verdades favoritas por la realización de series elaboradamente encadenadas de experiementos de laboratorio o por largas caminatas en el desierto. Estamos todos de pie alrededor de nuestros respectivos datos, mirando en casi las mismas direcciones, buscando casi las mismas cosas. Las disputas por prioridades pueden tener sentido en algunas disciplinas, pero en filosofía tienden a asumir la apariencia de disputas entre marineros sobre quién recibe el reconocimiento por ser el primero en observar que se ha levantado una brisa.

En el caso de estudio podría decirse que todo vuelve a Quine o que Sellars

[5] Para una hermenéutica ecuménica heroica, esta atracción de Heiddeger hacia el redil quineano sólo es emulada por el estudio de Wheeler (1986) de Derrida, Quine, Dennett y Davidson.

merece el reconocimiento o, para tocar una melodía tradicional, que todo no es más que una serie de notas al pie de página sobre Platón. Desde una posición de ventaja, parece que hubiera una migración gradual de teóricos hacia el Ecuador, que toman en serio lo que Quine denomina el "modismo dramático" de atribución intencional, pero no con ***demasiada*** seriedad, tratándola siempre como una "cubierta heurística" (Dennett, 1969) o una "actitud" (Dennett, 1971). Desde esta posición de ventaja podría inclusive parecer que si existiera una explicación intencional evidente de esta migración: los filósofos, al ser sistemas intencionales aproximadamente racionales, se están convenciendo gradualmente de que Dennett tiene razón. Pero ésa es, sin duda, una ilusión de perspectiva.

Bibliografía

Abbott, E. A. (1962): *Flatland: A Romance in Many Dimensions*, Oxford, Blackwell. (Publicado originalmente en 1884.)

Ackermann, R. (1972): "Opacity in Belief Structures", *Journal of Philosophy*, LXIX, págs. 55-67.

Akins, K. A. (1986): "On Piranhas, Narcissism, and Mental Representation", CCM-86-2, Center for Cognitive Studies, Tufts University.

Akins, K. A. (inédito): "Information and Organisms: or Why Nature Doesn't Build Epistemic Engines", disertación doctoral, University of Michigan, Ann Arbor, 1987.

Amundson, R. (inédito): "Doctor Dennett and Doctor Pangloss".

Anderson, A. R. y Belnap, N. (1974): *Entailment: The Logic of Relevance and Necessity*, Princeton, Princeton University Press.

Anscombe, E. (1957): *Intention*, Oxford, Blackwell.

Aquila, R. E. (1977): *Intentionality: A Study of Mental Acts*, University Park, Pennsylvania State University Press.

Barash, D. P. (1976): "Male Response to Apparent Female Adultery in the Mountain Bluebird: An Evolutionary Interpretation", *American Naturalist*, 110, págs. 1097-1101.

Beatty, J. (1980): "Optimal-design Models and the Strategy of Model Building in Evolutionary Biology", *Philosophy of Science*, 47, págs. 532-61.

Bechtel, W. (1985): "Realism, Reason, and the Intentional Stance", *Cognitive Science*, 9, págs. 473-97.

Bennett, J. (1964): *Rationality*, Londres, Routledge and Kegan Paul.

Bennett, J. (1976): *Linguistic Behavior*, Cambridge, Cambridge University Press.

Bennett, J. (1983): "Cognitive Ethology: Theory or Poetry?" (comentario sobre Dennett 1983a), *Behavioral and Brain Sciences*, 6, págs. 356-58.

Berliner, H. y Ebling, C. (1986): "The SUPREM Architecture: a new Intelligent Paradigm", *Artificial Intelligence*, 28, págs. 3-8.

Blackburn, S. (1979): "Thought and Things", *Aristotelian Society Supplementary Volume*, LIII, págs. 23-42.

Block, N. (1978): "Troubles with Functionalism", en C. W. Savage, comp., *Per-*

ception and Cognition: Issues in the Foundations of Psychology, Minneapolis, University of Minnesota Press. (Reimpreso en Block 1980, vol. 1.)
Block, N., comp. (1980): *Readings in the Philosophy of Psychology*, 2 vols., Cambridge, MA, Harvard University Press.
Boden, M. (1981): "The Case for a Cognitive Biology", en *Minds and Mechanisms: Philosophical Psychology and Computational Models*, Ithaca, Cornell University Press.
Boër, S. y Lycan, W. (1975): "Knowing Who", *Philosophical Studies*, 28, págs. 299-347.
Borges, J. L. (1962): "The Garden of Forking Paths", en D. A. Yates y J. E. Irby, comps., *Labyrinths: Selected Stories and Other Writings*, Nueva York, New Directions. [Original en castellano: *El jardín de senderos que se bifurcan*, Buenos Aires, Sur, 1941.]
Braitenberg, V. (1984): *Vehicles: Experiments in Synthetic Psychology*, Cambridge, MA, The MIT PRess/A. Bradford Book.
Burge, T. (1977): "Belief De Re", *The Journal of Philosophy*, 74, nº 6, págs. 338-62.
Burge, T. (1978): "Belief and Synonymy", *Journal of Philosophy*, 75, págs. 119-38.
Burge, T. (1979): "Individualism and the Mental", *Midwest Studies in Philosophy*, IV, págs. 73-121.
Burge, T. (1986): "Individualism and Psychology", *The Philosophical Review*, XCV, nº 1, págs. 3-46.
Byrne, R. y Whiten, A., comps. (en preparación): *Social Expertise and the Evolution of Intellect: Evidence from Monkeys, Apes, and Humans*, Oxford, Oxford University Press.
Cain, A. J. (1964): "The Perfection of Animals", *Viewpoints in Biology*, 3, págs. 37-63.
Campbell, D. T. (1973): "Evolutionary Epistemology", en Paul Schilpp, comp., *The Philosophy of Karl Popper*, La Salle, IL, Open Court Press.
Campbell, D. T. (1977): "Descriptive Epistemology: Psychological, Sociological, and Evolutionary", William James Lectures, Harvard University.
Cargile, J. (1970): "A Note on 'Iterated Knowings'", *Analysis*, 30, págs. 151-55.
Castaneda, H.-N. (1966): " 'He': A Study in the Logic of Self-Consciousness", *Ratio*, 8, págs. 130-57.
Castaneda, H.-N. (1967): "Indicators and Quasi-Indicators", *American Philosophical Quarterly*, 4, págs. 85-100.
Castaneda, H.-N (1968): "On the Logic of Attributions of Self-Knowledge to Others", *Journal of Philosophy*, LXV, págs. 439-56.
Cohen, L. B.; DeLoache, J. S. y Strauss, M. S. (1979): "Infant Visual Perception", en J. D. Osofsky, comp., *Handbook of Infant Development*, Nueva York, Wiley, págs. 416-19.
Cohen, L. J. (1981): "Can Human Rationality be Experimentally Demonstrated?", *Behavioral and Brain Sciences*, 4, págs. 317-70.
Charniak, E. (1974): "Toward a Model of Children's Story Comprehension", disertación doctoral inédita, MIT, y MIT AI Lab Report 266.
Cheney, D. y Seyfarth, R. (1982): "Recognition of Individuals Within and Between Groups of Free-Ranging Vervet Monkeys", *American Zoology*, 22, págs. 519-29.

Cheney, D. y Seyfarth, R. (1985): "Social and Non-social Knowledge in Vervet Monkeys", *Philosophical Transactions of the Royal Society of London*, B 308, págs. 187-201.
Cherniak, C. (1981): "Minimal Rationality", *Mind*, 90, págs. 161-83.
Cherniak, C. (1983): "Rationality and the Structure of Memory", *Synthese*, 57, págs. 163-86.
Cherniak, C. (1986): *Minimal Rationality*, Cambridge, MA, The MIT Press/A Bradford Book.
Chisholm, R. (1956): "Sentences About Believing", *Aristotelian Society Proceedings*, 56, págs. 125-48.
Chisholm, R. (1957): *Perceiving: A Philosophical Study*, Ithaca, Cornell University Press.
Chisholm, R. (1966): "On Some Psychological Concepts and the 'Logic' of Intentionality", en H. N. Castaneda, comp., *Intentionality, Minds, and Perception*, Detroit, Wayne State University Press.
Chisholm, R. y Sellars, W. (1958): "Intentionality and the Mental", en H. Feigl, M. Scriven y G. Maxwell, comps., *Concepts, Theories and the Mind-Body Problem*, Minnesota Studies in Philosophy of Science, II. Minneapolis, University of Minnesota Press.
Chomsky, N. (1959): "Review of B. F. Skinner's *Verbal Behavior*", *Language*, 35, págs. 26-58. (Reimpreso en Block 1980, vol. 1)
Chomsky, N. (1980a): *Rules and Representations*, Nueva York, Columbia University Press. [Hay versión castellana: *Reglas y representaciones*, México, Fondo de Cultura Económica.]
Chomsky, N. (1980b): "Rules and Representations", *Behavioral and Brain Sciences*, 3, págs. 1-61.
Churchland, P. M. (1979): *Scientific Realism and the Plasticity of Mind*, Cambridge, Cambridge University Press.
Churchland, P. M. (1981): "Eliminative Materialism and the Propositional Attitudes", *Journal of Philosophy*, 78, págs. 67-90.
Churchland, P. M. (1984): *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind*, Cambridge, MA, The MIT Press/A Bradford Book.
Churchland, P. S. (1980): "Language, Thought, and Information Processing", *Nous*, 14, págs. 147-70.
Churchland, P. S. (1986): *Neurophilosophy: Toward a Unified Theory of Mind/Brain*, Cambridge, MA, The MIT Press/A Bradford Book.
Churchland, P. S. y Churchland, P. M. (1981): "Stalking the Wild Epistemic Engine", *Nous*, págs. 5-18.
Dahlbom, B. (1985): "Dennett on Cognitive Ethology: a Broader View" (comentario sobre Dennett 1983a), *Behavioral and Brain Sciences*, 8, págs. 760-61.
Darmstadter, H. (1971): "Consistency of Belief", *Journal of Philosophy*, 68, págs. 301-10.
Davidson, D. (1967): "Truth and Meaning", *Synthese*, XVII, págs. 304-23.
Davidson, D. (1969): "How is Weakness of the Will Possible?", en J. Feinberg, comp., *Moral Concepts*, Oxford, Oxford University Press.
Davidson, D. (1970): "Mental Events", en L. Foster y J. Swanson, comps., *Experience and Theory*, Amherst, University of Massachusetts Press.

Davidson, D. (1973): "Radical Interpretation", *Dialectica*, 27, págs. 313-28. (Reimpreso en D. Davidson, *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press, 1985).
Davidson, D. (1974a): "Belief and the Basis of Meaning", *Synthese*, 27, págs. 309-23.
Davidson, D. (1974b): "On the Very Idea of a Conceptual Scheme", *Proceedings and Addresses of the American Philosophical Association*, 47, págs. 5-20.
Davidson, D. (1975): "Thought and Talk", en *Mind and Language: Wolfson College Lectures, 1974*, Oxford, Clarendon Press, págs. 7-23.
Davidson, D. (1985): *Inquiries into Truth and Interpretation*. Oxford, Clarendon Press. [Hay versión castellana: *Sobre la verdad y la interpretación*, Barcelona, Gedisa, 1990.]
Dawkins, R. (1976): *The Selfish Gene*, Oxford, Oxford University Press.
Dawkins, R. (1980): "Good Strategy or Evolutionarily Stable Strategy?", en G. W. Barlow y J. Silverberg, comps., *Sociobiology: Beyond Nature/Nurture?*, A.A.A.S. Selected Symposium, Boulder, CO, Westview Press.
Dawkins, R. (1982): *The Extended Phenotype*, San Francisco, Freeman.
Dawkins, R. (1986): *The Blind Watchmaker*, Essex, Longman Scientific and Technical.
Dennett, D. C. (1969): *Content and Consciousness*, Londres, Routledge and Kegan Paul.
Dennett, D. C. (1971): "Intentional Systems", *Journal of Philosophy*, 8, págs. 87-106. (Reimpreso en Dennett 1978a.)
Dennett, D. C. (1973): "Mechanism and Responsibility", en T. Honderich, comp., *Essays on Freedom of Action*, Londres, Routledge and Kegan Paul. (Reimpreso en Dennett 1978a.)
Dennett, D. C. (1974a): "Why the Law of Effect Will Not Go Away", *Journal of the Theory of Social Behavior*, 5, págs. 169-87. (Reimpreso en Dennett 1978a.)
Dennett, D. C. (1974b): "Comment on Wilfrid Sellars", *Synthese*, 27, págs. 439-44.
Dennett, D. C. (1975): "Brain Writing and Mind Reading", en K. Gunderson, comp., *Language, Mind, and Meaning*, Minnesota Studies in Philosophy of Science, VII, Minneapolis, University of Minnesota Press.
Dennett, D. C. (1976): "Conditions of Personhood", en A. Rorty, comp., *The Identities of Persons*, Berkeley, University of California Press. (Reimpreso en Dennett 1978a.)
Dennett, D. C. (1978a): *Brainstorms: Philosophical Essays on Mind and Psychology*, Montgomery, VT, Bradford Books.
Dennett, D. C. (1978b): "Beliefs About Beliefs" (comentario sobre Premack y Woodruff 1978), *Behavioral and Brain Sciences*, 1, págs. 568-70.
Dennett, D. C. (1978c): "Current Issues in the Philosophy of Mind", *American Philosophical Quarterly*, 15, págs. 249-61.
Dennett, D. C. (1978d): "Why Not the Whole Iguana?", (comentario sobre Pylyshyn 1978), *Behavioral and Brain Sciences*, 1, págs. 103-4.
Dennett, D. C. (1980a): "Passing the Buck to Biology" (comentario sobre Chomsky 1980b), *Behavioral and Brain Sciences*, 3, pág. 19.
Dennett, D. C. (1980b): "The Milk of Human Intentionality" (comentario sobre Searle 1980b), *Behavioral and Brain Sciences*, 3, págs. 428-30.

Dennett, D. C. (1980c): "Reply to Stich", *Philosophical Books*, 21, págs. 65-76.
Dennett, D. C. (1982a): "Comment on Rorty", *Synthese*, 53, págs. 349-56.
Dennett, D. C. (1982b): "How to Study Human Consciousness Empirically: or, Nothing Comes to Mind", *Synthese*, 53, págs. 159-80.
Dennett, D. C. (1982c): "The Myth of the Computer: An Exchange", *The New York Review of Books*, 24 de junio, págs. 56-57.
Dennett, D. C. (1982d): "Why Do We Think What We Do About Why We Think What We Do?", *Cognition*, 12, págs. 219-27.
Dennett, D. C. (1983a): "Intentional Systems in Cognitive Ethology: The 'Panglossian Paradigm' Defended", *Behavioral and Brain Sciences*, 6, págs. 343-90. (Reimpreso como capítulo 7 en este libro.)
Dennett, D. C. (1983b): "Artificial Intelligence and the Strategies of Psychological Investigation" (entrevista) en J. Miller, comp., *States of Mind*, Londres, BBC Publications.
Dennett, D. C. (1984a): "Carving the Mind at Its Joints" (crítica de Fodor 1983), *Contemporary Psychology*, 29, págs. 285-86.
Dennett, D. C. (1984b): "Computer Models and the Mind—a View from the East Pole", *Times Literary Supplement*, 14 de Diciembre de 1984, págs. 1453-54. (Este es un borrador anterior e incompleto de Dennett 1986e).
Dennett, D. C. (1984c): "Cognitive Wheels: the Frame Problem of AI", en C. Hookway, comp., *Minds, Machines and Evolution*, Cambridge, Cambridge University Press.
Dennett, D. C. (1984d): *Elbow Room: The Varieties of Free Will Worth Wanting*, Cambridge, MA, The MIT Press/A Bradford Book. [Hay versión castellana en preparación, Barcelona, Gedisa, 1991.]
Dennett, D. C. (1984e): "The Role of the Computer Metaphor in Understanding the Mind", en H. Pagels, comp., *Computer Culture: the Scientific, Intellectual, and Social Impact of the Computer*, Annals of the New York Academy of Sciences, vol. 426, págs. 266-75.
Dennett, D. C. (1985a): "Can Machines Think?", en M. Shafto, comp., *How We Know*, San Francisco, Harper and Row.
Dennett, D. C. (1985b): "When does the Intentional Stance Work?", *Behavioral and Brain Sciences*, 8, págs. 758-66.
Dennett, D. C. (1985c): "Why Believe in Belief?", (crítica de Stich 1983), *Contemporary Psychology*, vol. 30, pág. 949.
Dennett, D. C. (1986a): "Engineering's Baby" (comentario sobre Sayre 1986), *Behavioral and Brain Sciences*, 9, págs. 141-42.
Dennett, D. C. (1986b): "Information, Technology and the Virtues of Ignorance", *Daedalus*, 115, págs. 135-53.
Dennett, D. C. (1986c): "Is There an Autonomous 'Knowledge Level'?", (comentario sobre Newell, en el mismo volumen), en Z. Pylyshyn y W. Demopoulos, comps., *Meaning and Cognitive Structure: Issues in the Computational Theory of Mind*, Norwood, NJ, Ablex.
Dennett, D. C. (1986d): "Julian Jaynes' Software Archeology", *Canadian Psychology*, 27, págs. 149-54.
Dennett, D. C. (1986e): "The Logical Geography of Computational Approaches: a View from the East Pole", en R. Harnish y M. Brand, comps., *The Representation of Knowledge and Belief*, Tucson, University of Arizona Press.

Dennett, D. C. (próximo a aparecer a): "A Route to Intelligence: Oversimplify and Self-monitor", en J. Khalfa, comp., *Can Intelligence be Explained?*, Oxford, Oxford University Press.
Dennett, D. C. (próximo a aparecer b): "Cognitive Ethology: Hunting for Bargains or a Wild Goose Chase?", en D. McFarland, comp., *The Explanation of Goal-seeking Behaviour*, Oxford, Oxford University Press.
Dennett, D. C. (próximo a aparecer c): "Out of the Armchair and Into the Field", *Poetics Today*, Israel.
Dennett, D. C. (próximo a aparecer d): "Quining Qualia", en A. Marcel y E. Bisiach, comps., *Consciousness in Contemporary Science*, Oxford, Oxford University Press.
Dennett, D. C. (próximo a aparecer e): "The Moral First Aid Manual", 1986 Tanner Lecture, University of Michigan.
Dennett, D. C. (próximo a aparecer f): "The Myth of Original Intentionality", en W. Newton Smith y R. Viale, comps., *Modelling the Mind*, Oxford, Oxford University Press.
Dennett, D. C. (próximo a aparecer g): "The Self as the Center of Narrative Gravity", en P. COle, D. Johnson y F. Kessel, comps. *Consciousness and Self*, Nueva York, Praeger.
Dennett, D. C. y Haugeland, J. (1987): "Intentionality", en R. Gregory, comp., *The Oxford Companion to Mind*, Oxford, Oxford University Press.
de Sousa, R. (1971): "How to Give a Piece of Your Mind: Or, the Logic of Belief and Assent", *Review of Metaphysics*, 25, págs. 52-79.
de Sousa, R. (1979): "The Rationality of Emotion", *Dialogue*, 18, págs. 41-63.
Dewdney, A. K. (1984): *The Planiverse*, Nueva York, Poseidon.
Dobzhansky, T. (1956): "What is an Adaptive Trait?", *American Naturalist*, 90, págs. 337-47.
Donnellan, K. (1966): "Reference and Definite Descriptions", *Philosophical Review*, 75, págs. 281-304.
Donnellan, K. (1968): "Putting Humpty-Dumpty Together Again", *Philosophical Review*, 77, págs. 203-15.
Donnellan, K. (1970): "Proper Names and Identifying Descriptions", *Synthese*, 21, págs. 335-58.
Donnellan, K. (1974): "Speaking of Nothing", *Philosophical Review*, 83, págs. 3-31.
Dover Wilson, J. (1951): *What Happens in Hamlet*, 3ª ed., Cambridge, Cambridge University Press.
Dretske, F. (1981): *Knowledge and the Flow of Information*, Cambridge, MA, The MIT Press/A Bradford Book.
Dretske, F. (1985): "Machines and the Mental", Western Division APA Presidential Address, 26 de abril de 1985 (impreso en *Proceedings and Addresses of the APA* [1985] vol. 59, págs. 23-33.)
Dretske, F. (1986): "Misrepresentation", en R. Bogdan, comp., *Belief*, Oxford, Oxford University Press.
Dummett, M. (1973): *Frege: Philosophy of Language*, Londres, Duckworth.
Dummett, M. (1975): "What is a Theory of Meaning?", en S. Guttenplan, comp., *Mind and Language*, Oxford, Oxford University Press.

Enc, B. (1982): "Intentional States and Mechanical Devices", *Mind*, XCI, págs. 161-82.
Evans, G. (1973): "The Causal Theory of Names", *Aristotelian Society Supplementary Volume*, XLVII, págs. 187-208.
Evans, G. (1980): "Understanding Demonstratives", en H. Parret y J. Bouveresse, comps., *Meaning and Understanding*, Nueva York, Berlín, Walter de Gruyter.
Ewert, J.-P. (próximo a aparecer): "Neuroethology of Releasing Mechanisms: Prey-catching in Toads", *Behavioral and Brain Sciences*.
Fauconnier, G. (1985): *Mental Spaces*, Cambridge, MA, The MIT Press/A Bradford Book.
Feyerabend, P. (1978): *Science in a Free Society*, Londres, New Left Bank Publ.
Field, H. (1972): "Tarski's Theory of Truth", *Journal of Philosophy*, 69, págs. 347-74.
Field, H. (1977): "Logic, Meaning and Conceptual Role", *Journal of Philosophy*, 74, págs. 379-409.
Field, H. (1978): "Mental Representation", *Erkenntnis*, 13, págs. 9-61.
Fodor, J. (1968a): *Psychological Explanation: An Introduction to the Philosophy of Psychology*, Nueva York. Randolm House. [Hay versión castellana: *La explicación psicológica*, Cátedra, Madrid.]
Fodor, J. (1968b): "The Appeal to Tacit Knowledge in Psychological Explanation", *Journal of Philosophy*, 65, págs. 627-40.
Fodor, J. (1975): *The Language of Thought*, Hassocks, Sussex, Harvester Press; Scranton, PA, Crowell. [Hay versión castellana: *El lenguaje del pensamiento*, Alianza psicológica, Madrid.]
Fodor, J. (1980): "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology", *Behavioral and Brain Sciences*, 3, págs. 63-110. (Reimpreso en Fodor 1981a.)
Fodor, J. (1981a): *Representations*, Cambridge, MA, The MIT Press/A Bradford Book.
Fodor, J. (1981b): "Three Cheers for Propositional Attitudes", en Fodor 1981a.
Fodor, J. (1981c): "Tom Swift and his Procedural Grandmother", en Fodor 1981a.
Fodor, J. (1983): *The Modularity of Mind*, Cambridge, MA, The MIT Press/A Bradford Book. [Hay versión castellana: *La modularidad de la mente*, Morata, Madrid, 1986.]
Fodor, J. (1985): "Fodor's Guide to Mental Representation", *Mind*, XCIV, págs. 76-100.
Fodor, J. (1986): "Why Paramecia Don't Have Mental Representations", en *Midwest Studies in Philosophy*, X, págs. 3-23.
Fodor, J. (1987): *Psychosemantics*, Cambridge, MA, The MIT Press/A Bradford Book.
Frege, G. (1956): "The Thought: A Logical Inquiry", traducción de A. M. y M. Quinton, *Mind*, LXV, págs. 289-311. (Reimpreso en P. F. Strawson, comp., *Philosophical Logic*, Oxford, Oxford University Press, 1967.)
Friedman, M. (1981): "Theoretical Explanation", en R. Healy, comp., *Reduction, Time and Reality*, Cambridge, Cambridge University Press, págs. 2-31.
Gardner, H. (1975): *The Shattered Mind: The Person After Brain Damage*, Nueva York, Knopf.

Gardner, M. (1970): "Mathematical Games", *Scientific American*, 223, nº 4, págs. 120-23.
Gazzaniga, M. (1985): *The Social Brain: Discovering the Networks of the Mind*, Nueva York, Basic Books.
Gazzaniga, M. y Ledoux, J. E. (1978): *The Integrated Mind*, Nueva York, Plenum Press.
Geach, P. (1957): *Mental Acts*, Londres Routledge and Kegan Paul.
Ghiselin, M. T. (1983): "Lloyd Morgan's Canon in Evolutionary Context", (comentario sobre Dennett 1983a), *Behavioral and Brain Sciences*, 6, págs. 362-63.
Gibson, E. (1969): *Principles of Perceptual Learning and Development* (Century Psychology Series), Nueva York, Appleton-Century-Crofts.
Goldman, A. (1986): *Epistemology and Cognition*, Cambridge, MA, Harvard University Press.
Goodman, N. (1961): "About", *Mind*, 71, págs. 1-24.
Goodman, N. (1978): *Ways of Worldmaking*, Indianapolis, Hackett.
Goren, C. G.; Sorty, M. y Wu, P. Y. K. (1975): "Visual Following and Pattern Discrimination of Face-like Stimuli by Newborn Infants", *Pediatrics*, 56, págs. 544-49.
Gould, J. L. y Gould C. G. (1982): "The Insect Mind: Physics or Metaphysics?", en D. R. Griffin, comp., *Animal Mind—Human Mind*, Berlín, Springer-Verlag.
Gould, S. J. (1977): *Ever Since Darwin*, Nueva York, W. W. Norton and Co. [Hay versión castellana: *Desde Darwin (reflexiones sobre Historia Natural)*, Barcelona, Blume (Hermann), 1983.]
Gould, S. J. (1980): *The Panda's Thumb*. Nueva York, W. W. Norton and Co. [Hay versión castellana: *El pulgar del panda*, Orbis, Madrid, 1983.]
Gould, S. J. y Lewontin, R. (1979): "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme", *Proceedings of the Royal Society*, B205, págs. 581-98.
Grandy, R. (1973): "Reference, Meaning and Belief", *Journal of Philosophy*, 70, págs. 439-52.
Gregory, R. (1977): *Eye and Brain*, 3ª ed., Londres, Weidenfeld and Nicolson.
Grice, H. P. (1957): "Meaning", *Philosophical Review*, 66, págs. 377-88.
Grice, H. P. (1969): "Utterer's Meaning and Intentions", *Philosophical Review*, 78, págs. 147-77.
Griffin, D. R., comp. (1982): *Animal Mind—Human Mind*, Berlín, Springer-Verlag.
Hampshire, S. (1975): *Freedom of the Individual*, edición ampliada, Princeton, Princeton University Press.
Harman, G. (1968): "Three Levels of Meaning", *Journal of Philosophy*, LXV, págs. 590-602.
Harman, G. (1973): *Thought*, Princeton, Princeton University Press.
Harman, G. (1977): "How to Use Propositions", *American Philosophical Quarterly*, 14, págs. 173-76.
Harman, G. (1983): "Conceptual Role Semantics", *Notre Dame Journal of Formal Logic*, 28, págs. 242-56.
Harman, G. (1986): "Wide Functionalism", en R. Harnish y M. Brand, comps.,

The Representation of Knowledge and Belief, Tucson, University of Arizona Press.
Harman, G. (de próxima aparición): "(Nonsolipsistic) Conceptual Role Semantics" en E. Lepore, comp., *Semantics of Natural Language*, Nueva York, Academic Press.
Haugeland, J. (1981): *Mind Design*, Cambridge, MA, The MIT Press/A Bradford Book.
Haugeland, J. (1985): *Artificial Intelligence: The Very Idea*, Cambridge, MA, The MIT Press/A Bradford Book.
Haugeland, J. (inédito): "The Same Only Different".
Hayes, P. (1978): "Naive Physics I: The Ontology of Liquids", artículo de trabajo 35, Institut pour les Etudes Semantiques et Cognitives, Univ. de Genève.
Hayes, P. (1979): "The Naive Physics Manifesto", en D. Michie, comp., *Expert Systems in the Microelectronic Age*, Edimburgo, Edinburgh University Press.
Heyes, C. M. (de próxima aparición): "Cognisance of Consciousness in the Study of Animal Knowledge", en W. Callebaut y R. Pinxten, comps., *Evolutionary Epistemology: a Multiparadigm Approach*, Dordrecht, Reidel.
Hintikka, J. (1962): *Knowledge and Belief*, Ithaca, Cornell University Press. [Hay versión castellana: *Saber y creer. Una introducción a la lógica de las dos nociones*, Madrid, Tecnos, 1979.]
Hofstadter, D. (1979): *Gödel, Escher, Bach: An Eternal Golden Braid*, Nueva York, Basic Books.
Hofstadter, D. y Dennett, D. C. (1981): *The Mind's I: Fantasies and Reflections on Mind and Soul*, Nueva York, Basic Books.
Hornsby, J. (1977): "Singular Terms in Context of Propositional Attitude", *Mind*, LXXXVI, págs. 31-48.
House, W. (inédito): "Charity and the World According to the Speaker", disertación doctoral, University of Pittsburgh, 1980.
Humphrey, N. K. (1976): "The Social Function of Intellect", en P. P. G. Bateson y R. A. Hinde, comps., *Growing Points in Ethology*, Cambridge, Cambridge University Press.
Hunter, I. M. L. (1962): "An Exceptional Talent for Calculative Thinking", *British Journal of Psychology*, 53-54, págs. 243-58.
Israel, D. (inédito): "The Role of Propositional Objects of Belief in Action", CSLI Report, Center for the Study of Language and Information, Stanford University.
Ittleson, W. H. (1952): *The Ames Demonstrations in Perception*, Oxford, Oxford University Press.
Jackendoff, R. (1983): *Semantics and Cognition*, Cambridge, MA, The MIT Press/A Bradford Book.
Jackendoff, R. (1985): "Information is in the Mind of the Beholder", *Linguistics and Philosophy*, 8, págs. 23-33.
Jacob, F. (1977): "Evolution and Tinkering", *Science*, 196, págs. 1161-66.
Jeffrey, R. (1970): "Dracula Meets Wolfman: Acceptance vs. Partial Belief", en M. Swain, comp., *Induction, Acceptance and Rational Belief*, Dordrecht, Reidel.
Johnston, T. D. (1981): "Contrasting Approaches to a Theory Of Learning", *Behavioral and Brain Sciences*, 4, págs. 125-73.

Kahneman, D. (inédito): "Some Remarks on the Computer Metaphor".
Kahneman, D. y Tversky, A. (1983): "Choices, Values, and Frames", *American Psychologist*, 39, págs. 341-50.
Kaplan, D. (1968): "Quantifying In", *Synthese*, 19, págs. 178-214. (Reimpreso en *Words and Objections*, D. Davidson y J. Hintikka, comps. Dordrecht, Reidel, 1969.)
Kaplan, D. (1973): "Bob and Carol and Ted and Alice", en J. Hintikka, J. Moravcsik y P. Suppes, comps., *Approaches to Natural Language*, Dordrecht, Reidel.
Kaplan, D. (1978): "Dthat", en P. Cole, comp., *Syntax and Semantics*, Nueva York, Academic Press.
Kaplan, D. (1980): "Demonstratives", The John Locke Lectures, Oxford University.
Kitcher, P. (1984): "In Defense of Intentional Psychology", *Journal of Philosophy*, LXXI, págs. 89-106.
Kitcher, Ph. (1985): *Vaulting Ambition*, Cambridge, MA, The MIT Press.
Kitcher, Ph. (1987): "Why Not the Best?", en John Dupre, comp., *The Latest on the Best*. Cambridge, MA, The MIT Press/A Bradford Book.
Kiteley, M. (1968): "Of What We Think", *American Philosophical Quarterly*, 5, págs. 31-42.
Kripke, S. (1972): "Naming and Necessity", en D. Davidson y G. Harman, comps., *Semantics of Natural Language*, Dordrecht, Reidel.
Kripke, S. (1977): "Speaker's Reference and Semantic Reference", en P. French y otros, comps., *Midwest Studies in Philosophy*, II, Minneapolis, University of Minnesota Press, págs. 255-76.
Kripke, S. (1979): "A Puzzle About Belief", en A. Margolis, comp., *Meaning and Use*, Dordrecht, Reidel, págs. 239-83.
Kripke, S. (1982): *Wittgenstein on Rules and Private Language*, Cambridge, MA, Harvard University Press.
Lettvin, J. Y. y otros (1959): "What the Frog's Eye Tells the Frog's Brain", en *Proceedings of the Institute of Radio Engineers*, 1959, págs. 1940-51.
Levin, J. (de próxima aparición): "Must Reasons be Rational?", *Philosophy of Science*.
Lewin, R. (1980): "Evolutionary Theory Under Fire", *Science*, 210, págs. 881-87.
Lewis, D. (1974): "Radical Interpretation", *Synthese*, 23, págs. 331-44.
Lewis, D. (1978): "Truth in Fiction", *American Philosophical Quarterly*, 15, págs. 37-46.
Lewis, D. (1979): "Attitudes *De Dicto* and *De Se*", *Philosophical Review*, 78, págs. 513-43.
Lewontin, R. (1961): "Evolution and the Theory of Games", *Journal of Theoretical Biology*, 1, págs. 328-403.
Lewontin, R. (1978a): "Adaptation", *Scientific American*, 293, nº 3 (Septiembre), págs. 213-30.
Lewontin, R. (1978b): "Fitness, Survival, and Optimality", en D. H. Horn, R. Mitchell y G. R. Stairs, comps., *Analysis of Ecological Systems*, Cincinnati, Ohio State University Press.
Lewontin, R. (1979): "Sociobiology as an Adaptionist Paradigm", *Behavioral Science*, 24, págs. 5-14.

Lewontin, R. (1981): "The Inferiority Complex", *The New York Review of Books*, 22 de octubre, págs. 12-16.
Livingston, R. (1978): *Sensory Processing, Perception and Behavior*, Nueva York, Raven Press.
Lloyd, M. y Dybas, H. S. (1966): "The Periodical Cicada Problem", *Evolution*, 20, págs. 132-49.
Loar, B. (1972): "Reference and Propositional Attitudes", *Philosophical Review*, 80, págs. 43-62.
Loar, B. (1981): *Mind and Meaning*, Cambridge, Cambridge University Press.
Loar, B. (de próxima aparición): "Social Content and Psychological Content", en D. Merrill y R. Grimm, comps., *Content of Thought*, Tucson, University of Arizona Press.
Lycan, W. (1974): "Mental States and Putnam's Functionalist Hypothesis", *Australasian Journal of Philosophy*, 52, págs. 48-62.
Lycan, W. (1981a): "Form, Function and Feel", *Journal of Philosophy*, LXXVIII, págs. 24-49.
Lycan, W. (1981b): "Psychological Laws", *Philosophical Topics*, 12, págs. 9-38. (impreso en J. I. Biro y R. W. Shahan, comps., *Mind, Brain, and Function: Essays in Philosophy of Mind*, Norman, University of Oklahoma Press, 1982).
MacKenzie, A. W. (inédito): "Intentionality-One: Intentionality-Two", presentado en las reuniones de la Canadian Philosophical Association, 1978.
Marcus, R. B. (1983): "Rationality and Believing the Impossible", *Journal of Philosophy*, LXXX, págs. 321-37.
Marks, C. (1980): *Commissurotomy, Consciousness and the Unity of Mind*, Montgomery, VT, Bradford Books.
Marler, P.; Dufty, A. y Pickert, R. (1986a): "Vocal Communication in the Domestic Chicken I: Does a Sender Communicate Information about the Quality of a Food Referent to a Receiver?", *Animal Behavior*, 34, págs. 188-93.
Marler, P.; Dufty, A. y Pickert, R. (1986b): "Voice Communication in the Domestic Chicken II; Is the Sender Sensitive to the Presence and Nature of the Receiver?", *Animal Behavior*, 34, págs. 194-98.
Marr, D. (1982): *Vision*, Cambridge, MA, The MIT Press. [Hay versión castellana; *La visión*, Alianza, Madrid].
Maurer, D. y Barrera, M. (1981): "Infant's Perception of Natural and Distorted Arrangements of a Schematic Face", *Child Development*, 52, págs. 196-202.
Maynard Smith, J. (1972): *On Evolution*, Edimburgo, Edinburgh University Press.
Maynard Smith, J. (1974): "The Theory of Games and the Evolution of Animal Conflict", *Journal of Theoretical Biology*, 49, págs. 209-21.
Maynard Smith, J. (1978): "Optimization Theory in Evolution", *Annual Review of Ecology and Systematics*, 9, págs. 31-56.
Maynard Smith, J. (1983): "Adaptationism and Satisficing" (comentario sobre Dennett 1983a), *Behavioral and Brain Sciences*, 6, págs. 370-71.
Mayr, E. (1983): "How to Carry out the Adaptationist Program", *American Naturalist*, 121, págs. 324-34.
McCarthy, J. (1960): "Programs with Common Sense", en D. V. Blake y A. M. Uttley, comps., *Proceedings of the Symposium on Mechanization of Thought Processes*, National Physical Laboratory, Teddington, Inglaterra, H. M. Stationery Office, págs. 75-91. (Reimpreso en Minsky 1968.).

McCarthy, J. (1979): "Ascribing Mental Qualities to Machines", en M. Ringle, comp., *Philosophical Perspectives in Artificial Intelligence*, Atlantic Highlands, NJ, Humanities Press.
McCarthy, J. Hayes P. (1969): "Some Philosophical Problems from the Standpoint of Artificial Intelligence", en B. Meltzer y D. Michie, comps., *Machine Intelligence*, Edimburgo, Edinburgh University Press.
Mc Clelland, J. y Rumelhart, D., comps. (1986): *Parallel Distributed Processing Explorations in the Microstructures of Cognition*, 2 vols., Cambridge, MA, The MIT Press/A Bradford Book.
McDowell, J. (1977): "On The Sense and Reference of a Proper Name", *Mind*, LXXXVI, págs. 159-85.
McFarland, D. (1984): *Animal Behavior*, Menlo Park, CA, Benjamin-Cummings Publ.
Miller, J., comp. (1983): *States of Mind*, Londres, BBC Publications.
Millikan, R. (1984): *Language, Thought and Other Biological Categories*, Cambridge, MA, The MIT Press/A Bradford Book.
Millikan, R. (1986): "Thoughts Without Laws: Cognitive Science Without Content", *Philosophical Review*, XCV, págs. 47-80.
Millikan, R. (inédito): "Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox".
Minsky, M., comp. (1968): *Semantic Information Processing*, Cambridge, MA, The MIT Press.
Monod, J. (1971): *Chance and Necessity*, Nueva York, Knopf Press. (Originalmente publicada en Francia como *Le Hasard et la Necessité*, París, Editions du Seuil, 1970.) [Hay versión castellana: *El azar y la necesidad*, Orbis, Barcelona, 1985.]
Morton, A. (1975): "Because He Thought He had Insulted Him", *Journal of Philosophy*, LXXII, págs. 5-15.
Nagel, T. (1979): *Mortal Questions*, Cambridge, Cambridge University Press.
Nagel, T. (1986): *The View From Nowhere*, Oxford, Oxford University Press.
Neisser, U. (1976): *Cognition and Reality*, San Francisco, Freeman. [Hay versión castellana: *Procesos cognitivos y realidad*, Madrid, Marova, 1981.]
Nelson, R. J. (1978): "Objects of Occasion Beliefs", *Synthese*, 39, págs. 105-40.
Newell, A. (1982): "The Knowledge Level", *Artificial Intelligence*, 18, págs. 81-132.
Nisbett, R. E. y Ross, L. D. (1980): *Human Inference: Strategy and Shortcomings*, Englewood Cliffs, Prentice Hall.
Nisbett, R. E. y Wilson, T. DeC. (1977): "Telling More than We Know: Verbal Reports on Mental Processes", *Psychological Review*", 84, págs. 231-59.
Nozick, R. (1981): *Philosophical Explanations*, Cambridge, MA, Harvard University Press.
Oster, G. F. y Wilson, E. O. (1978): *Caste and Ecology in the Social Insects*, Princeton, Princeton University Press.
Parfit, D. (1984): *Reasons and Persons*, Oxford, Oxford University Press.
Perry, J. (1977): "Frege on Demonstratives", *Philosophical Review*, 86, págs. 474-97.
Perry, J. (1979): "The Problem of the Essential Indexical", *Nous*, 13, págs. 3-21.
Popper, K. y Eccles, J. (1977): *The Self and its Brain*, Berlín, Springer-International. [Hay versión castellana: *El yo y su cerebro*, Barcelona, Labor, 1982.]

Powers, L. (1978): "Knowledge by Deduction", *Philosophical Review*, LXXXVII, págs. 337-71.
Premack, A. (1983): "The Codes of Man and Beasts", *Behavioral and Brain Sciences*, 6, págs. 125-68.
Premack, A. (1986): *Cavagai! Or the Future History of the Animal Language Controversy*, Cambridge, MA, The MIT Press/A Bradford Book.
Premack, A. y Woodruff, G. (1978): "Does the Chimpanzee Have a Theory of Mind?", *Behavioral and Brain Sciences*, 1, págs. 515-26.
Prior, A. N. (1976): *Papers in Logic and Ethics*, comp. por P. Geach y A. Kenny, Londres, Duckworth.
Putnam, H. (1960): "Minds and Machines", en S. Hook, comp., *Dimensions of Mind*, Nueva York University Press. (Reimpreso en Putnam 1975b.)
Putnam, H. (1965): "Brains and Behavior", en J. Butler, comp. *Analytical Philosophy*, Oxford, Blackwell.
Putnam, H. (1974): "Comment on Wilfrid Sellars", *Synthese*, 27, págs. 445-55.
Putnam, H. (1975a): "The Meaning of Meaning", en Putnam 1975b.
Putnam, H. (1975b): *Mind, Language and Reality*, Philosophical Papers, II, Cambridge, Cambridge University Press.
Putnam, H. (1978): *Meaning and the Moral Sciences*, Londres, Rowledge and Kegan Paul.
Putnam, H. (1981): *Reason, Touth and History*, Cambridge, Cambridge University Press.
Putnam, H. (1983): "Computational Psychology and Interpretation Theory", en *Realism and Reason*, Philosophical Papers, III, Cambridge, Cambridge University Press.
Putman, H. (1986): "Information and the Mental", en E. Lepore, comp., *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford, Blackwell.
Pylyshyn, Z. (1978): "Computational Models and Empirical Constraint", *Behavioral and Brain Sciences*, 1, págs. 98-128.
Pylyshyn, Z. (1979): "Complexity and the Study of Artificial and Human Intelligence", en M. Ringle, comp., *Philosophical Perspectives in Artificial Intelligence*, Atlantic Highlands, NJ, Humanities Press.
Pylyshyn, Z. (1980): "Computation and Cognition: Issues in the Foundation of Cognitive Science", *Behavioral and Brain Sciences*, 3, págs. 111-32.
Pylyshyn, Z. (1984): *Computation and Cognition: Toward a Foundation for Cognitive Science*, Cambridge, MA, The MIT Press/A Bradford Book.
Quine, W. V. O. (1956): "Quantifiers and Propositional Attitudes", *Journal of Philosophy*, LIII, págs. 177-86. (Reimpreso en Quine, *The Ways of Paradox*, Nueva York, Random House, 1966.)
Quine, W. V. O. (1960): *Word and Object*, Cambridge, MA, The MIT Press.
Quine, W. V. O. (1969): "Propositional Objects", en *Ontological Relativity and Other Essays*, Nueva York, Columbia University Press, págs. 139-60. [Hay versión castellana: *La relatividad ontológica y otros ensayos*, Madrid, Tecnos, 1974.]
Quine, W. V. O. (1970): "On the Reasons for Indeterminacy of Translation", *Journal of Philosophy*, LXVII, págs. 178-83.
Ramachandran, V. S. (1985a): "Apparent Motion of Subjective Surfaces", *Perception*, 14, págs. 127-34.

Ramachandran, V. S. (1985b): Editorialista invitado en *Perception*, 14, págs. 97-103.
Raphael, B. (1976): *The Thinking Computer: Mind Inside Matter*, San Francisco, Freeman. [Hay versión castellana: *El computador pensante. Introducción a la informática para psicólogos y humanistas*, Madrid, Cátedra, 1984.]
Reichenbach, H. (1938): *Experience and Prediction*, Chicago, University of Chicago Press.
Ristau, C.A. (de próxima aparición): "Thinking Communicating, and Deceiving: Means to Master the Social Environment", en G. Greenberg y E. Tobach, comps., *Evolution of Social Behavior and Integrative Levels*, T. C. Schneirla Conference Series, Hillsdale, NJ, Erlbaum.
Ristau, C. A. (inédito): "Intentional Behavior by Birds?: The Case of the 'Injury Feigning' Plovers".
Roitblat, H. L. (1982): "The Meaning of Representation in Animal Memory", *Behavioral and Brain Sciences*, 5, págs. 352-406.
Rorty, R. (1979): *Philosophy and the Mirror of Nature*, Princeton, Princeton University Press. [Hay versión castellana: *La filosofía y el espejo de la naturaleza*, Madrid, Cátedra, 1983.]
Rorty, R. (1982): "Contemporary Philosophy of Mind", *Synthese*, 53, págs. 323-48.
Rosenberg, A. (1980): *Sociobiology and the Preemption of Social Science*, Baltimore, Johns Hopkins University Press.
Rosenberg, A. (1985): "Adaptationalist Imperatives and Panglossian Paradigms", en J. Fetzer, comp., *Sociobiology and Epistemology*, Dordrecht, Reidel.
Rosenberg, A. (1986a): "Intention and Action Among the Macromolecules", en N. Rescher, comp., *Current Issues in Teleology*, Lanham, NY, University Presses of America.
Rosenberg, A. (1986b): "Intentional Psychology and Evolutionary Biology (Part I: The Uneasy Analogy)", *Behaviorism*, 14, págs. 15-27.
Russell, B. (1905): "On Denoting", *Mind*, págs. 479-93. (Reimpreso en Russell, *Logic and Knowledge*, Londres, Allen and Unwin, 1958.) [Hay versión castellana: *Lógica y conocimiento*, Taurus, 1981, 2ª ed.]
Russell, B. (1959): *Mysticism and Logic*, Londres, Allen and Unwin. [Hay versión castellana: *Misticismo y lógica*, Buenos Aires, Paidós.]
Ryle, G. (1949): *The Concept of Mind*, Londres, Hutchinson. [Hay versión castellana: *El concepto de lo mental*, Buenos Aires, Paidós.]
Ryle, G. (1958): "A Puzzling Element in the Notion of Thinking", a British Academy Lecture. (Reimpreso en P. F. Strawson, comp., *Studies in the Philosophy of Thought and Action*, Oxford, Oxford University Press, 1968.)
Ryle, G. (1979): *On Thinking*, comp. por K. Kolenda, Totowa, NJ, Rowman and Littlefield.
Sacks, O. (1984): *A Leg to Stand On*, Nueva York, Summit Books.
Sacks, O. (1986): *The Man who Mistook His Wife for a Hat, and Other Clinical Tales*, Nueva York, Summit Books. [Hay versión castellana: *El hombre que confundió a su mujer con un sombrero*, Barcelona, Muchnik, 1987.]
Savage-Rumbaugh, S.; Rumbaugh, D. M. y Boysen, S. (1978): "Linguistically Mediated Tool Use and Exchange by Chimpanzees *(Pan troglodytes)*, *Behavioral and Brain Sciences*, 1, págs. 539-54.

Savage, C. W., comp. (1978): *Perception and Cognition: Issues in the Foundations of Psychology*, Minneapolis, University of Minnesota Press.
Sayre, K. (1986): "Intentionality and Information Processing: An Alternative Model for Cognitive Science", *Behavioral and Brain Sciences*, 9, págs. 121-60.
Schank, R. C. (1976): Informe de investigación 84, Yale University Department of Computer Science.
Schank, R. y Abelson, R. (1977): *Scripts, Plans, Goals and Understanding*, Hillside, NJ, Erlbaum.
Scheffler, I. (1963): *The Anatomy of Inquiry*, Nueva York, Knopf.
Schiffer, S. (1978): "The Basis of Reference", *Erkenntnis*, 13, págs. 171-206.
Schull, J. (de próxima aparición): "Evolution and Learning Analogies and Interactions", en E. Laszlo, comp. *The Evolutionary Paradigm: Transdisciplinary Studies*, Durham, Duke University Press.
Searle, J. (1979): "Referential and Attributive", *The Monist*, 62, págs. 190-308. (Reimpreso en Searle 1980a).
Searle J. (1980a): *Expression and Meaning*, Cambridge, Cambridge University Press.
Searle, J. (1980b): "Minds, Brains, and Programs", *Behavioral and Brain Sciences*, 3 págs. 417-58.
Searle, J. (1982): "The Myth of the Computer. An Exchange", *The New York Review of Books*, 24 de junio, págs. 56-57.
Searle, J. (1983): *Intentionality: An Essay in the Philosophy of Mind*, Cambridge, Cambridge University Press.
Searle, J. (1984): "Panel Discussion: Has Artificial Intelligence Research Illuminated Human Thinking?", en H. Pagels, comp., *Computer Culture: The Scientific, Intellectual, and Social Impact of the Computer*, Anales de la New York Academy of Sciences, vol. 426.
Searle, J. (1985): *Minds, Brains and Science*, Cambridge, MA, Harvard University Press. [Hay versión castellana: *Mentes, cerebros y ciencia*, Cátedra, Madrid.]
Searle, J. (de próxima aparición): "Turing the Chinese Room", en *Artificial Intelligence.*
Sejnowski, T. (de próxima aparición): "Computing With Connections" (crítica de W. D. Hillis, *The Connection Machine*, Cambridge, MA, The MIT Press, 1985), *Journal of Mathematical Psychology*.
Sejnowski, T. y Rosenberg, C. R. (1986): "NETtalk: A Parallel Network that Learns to Read Aloud", The Johns Hopkins University Electrical Engineering and Computer Science Technical Report JHU/EEC-86/01.
Sellars, W. (1954): "Some Reflections on Language Games", *Philosophy of Science*, 21, págs. 204-28. (reimpreso con modificaciones en Sellars 1963).
Sellars, W. (1956): "Empiricism and the Philosophy of Mind", en H. Feigl y M. Scriven, comps., *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, Minnesota Studies in the Philosophy of Science, I. Minneapolis, University of Minnesota Press. (Reimpreso en Sellars, 1963.)
Sellars, W. (1963): *Science, Perception and Reality*, Londres, Routledge and Kegan Paul.
Sellars, W. (1974): "Meaning as Functional Classification: a Perspective on the Relation of Syntax to Semantics", *Synthese*, 27, págs. 417-38.

Seyfarth, R.; Cheney, D. L. y Marler, P. (1980): "Monkey Responses to Three Different Alarm Calls: Evidence of Predator Classification and Semantic Communication", *Science*, 210, págs. 801-3.
Shaftz, M.; Wellman, H. y Silver, S. (1983): "The Acquisition of Mental Verbs: A Systematic Investigation of the First Reference to Mental States", *Cognition*, 14, págs. 301-21.
Shannon, C. (1949): *The Mathematical Theory of Communication*, Champaign-Urbana, University of Illinois Press.
Simmons, K. E. L. (1952): "The Nature of Predator Reactions of Breeding Birds", *Behaviour*, 4 págs. 101-76.
Simon, H. (1957): *Models of Man*, Nueva York, Wiley.
Simon, H. (1969): *The Sciences of the Artificial*, Cambridge, MA, The MIT Press. [Hay versión castellana: *Las ciencias de lo artificial*, Barcelona, Asesoría Técnica de Ediciones, 1979.]
Skinner, B. F. (1964): "Behaviorism at Fifty", en T. W. Wann, comp. *Behaviorism and Phenomenology: Contrasting Bases for Modern Psychology*, Chicago, University of Chicago Press.
Skinner, B. F. (1971): *Beyond Freedom and Dignity*, Nueva York, Knopf. [Hay versión castellana: *Más allá de la libertad y la dignidad*, Barcelona, Fontanella, 1972.]
Skutch, A. F. (1976): *Parent Birds and Their Young*, Austin, University of Texas Press.
Smith, S. B. (1983): *The Mental Calculators*, Nueva York, Columbia University Press.
Smolensky, P. (de próxima aparición): "Connectionist AI, Symbolic AI, and the Brain", *AI Review*.
Sober, E. (1981): "The Evolution of Rationality", *Synthese*, 46, págs. 95-120.
Sober, E. (1984): *The Nature of Selection*, Cambridge, MA, The MIT Press/A Bradford Book.
Sober, E. (1985): "Methodological Behaviorism, Evolution, and Game Theory", en James Fetzer, comp., *Sociobiology and Epistemology*, Dordrecht, Reidel.
Sordahl, T. A. (1981): "Sleight of Wing", *Natural History*, 90, págs. 43-49.
Sosa, E. (1970): "Propositional Attitudes *De Dicto* and *De Re*" *Journal of Philosophy*, 67, págs. 883-96.
Stabler, E. (1983): "How are Grammars Represented?", *Brain and Behavioral Sciences*, 6, págs. 391-422.
Stalnaker, R. (1976): "Propositions", en A. McKay y D. Merrill, comps., *Issues in the Philosophy of Language*, New Haven, Yale University Press.
Stalnaker, R. (1984): *Inquiry*, Cambridge, MA, The MIT Press/A Bradford Book.
Stalnaker, R. (inédito): "On What's in the Head".
Stich, S. (1978a): "Autnomous Psychology and the Belief-Desire Thesis". *The Monist*, 61, págs. 571-91.
Stich, S. (1978b): "Beliefs and Sub-Doxastic States", *Philosophy of Science*, 45, págs. 499-518.
Stich, S. (1980): "Headaches" (crítica de *Brainstorms*), *Philosophical Books*, XXI, págs. 65-76.
Stich, S. (1981): "Dennett on Intentional Systems", *Philosophical Topics*", 12, págs. 38-62.

Stich, S. (1982): "On the Ascription of Content", en A. Woodfield, comp., *Though and Content*, Oxford, Oxford University Press.
Stich, S. (1983): *From Folk Psychology to Cognitive Science. The Case Agains Belief*, Cambridge, MA, The MIT Press/A Bradford Book.
Stich, S. (1984): "Relativism, Rationality, and the Limits of Intentional Descrip tion", *Pacific Philosophical Quarterly*, 65, págs. 211-35.
Stich, S. y Nisbett, R. (1980): "Justification and the Psychology of Human Reason ing", *Philosophy of Science*, 47, págs. 188-202.
Stryer, L. (1981): *Biochemistry*, San Francisco, Freeman. [Hay versión castella na: *Bioquímica*, Barcelona, Reverté, 1982.]
Taylor, C. (1964): *The Explanation of Behaviour*, Londres, Routledge and Kegar Paul.
Thomason, R. (1986): "The Multiplicity of Belief and Desire", en E. Lepore, comp. *Truth and Interpretation: Perspectives on the Philosophy of Donald David son*, Oxford, Basil Blackwell.
Touretzky, D. S. e Hinton, G.E. (1985): "Symbols among the Neurons: Details o a Connectionist Inference Architecture", *Proceedings of the Ninth Interna cional Joint Conference on Artificial Intelligence*, Los Altos, Morgan Kauf man, págs. 238-43.
Trivers, R. L. (1971): "The Evolution of Reciprocal Altruism", *Quarterly Revieu of Biology*, 46, págs. 35-57.
Tversky, A. y Kahneman, D. (1974): "Judgement Under Uncertainty: Heuristic and Biases", *Science*, 185, págs. 499-518.
Ullian, J. y Goodman, N. (1977): "Truth About Jones", *Journal of Philosophy* LXXIV, págs. 317-38.
Vendler, Z. (1976): "Thinking of Individuals", *Nous*, 10, págs. 35-46.
Vendler, Z. (1981): "Reference and Introduction", *Philosophia*. (Reimpreso co modificaciones como el capítulo 4 de Vendler 1984.)
Vendler, Z. (1984): *The Matter of Minds*, Oxford, Clarendon Press.
Wallace J. (1972): "Belief and Satisfaction", *Nous*, 6, págs. 87-103.
Walton, K. (1978): "Fearing Fiction", *Journal of Philosophy*, 75. págs. 5-27.
Wason, P. y Johnson-Laird, P. (1972): *Psychology of Reasoning: Structure anc Content*, Londres, B. T. Batsford.
Weiskrantz, J. (1983): "Evidence and Scotomata", *Behavioral and Brain Sci ences*, 6, págs. 464-67.
Weizenfeld, J. (1977): "Surprise and Intentional Content" presentado en el III En cuentro Anual de la Sociedad de Filosofía y Psicología, Pittsburgh, marzo de 1977.
Wertheimer, R. (1974): "Philosophy on Humanity", en R. L. Perkins, comp., *Abor tion: Pro and Con*, Cambridge, MA, Schenkman.
Wheeler, S. C. (1986): "Indeterminacy of French Interpretation: Derrida and Da vidson", en E. Lepore, comp., *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford, Basil Blackwell.
Wilson, E. O.; Durlach, N. I. y Roth, L. M. (1958): "Chemical Releasers of Necro phoric Behavior in Ants", *Psyche*, 65, págs. 108-14.
Wilson, E. O. (1975): *Sociobiology: The New Synthesis*, Cambridge, MA, Harvar University Press. [Hay versión castellana: *Sociobiología*, Barcelona, Omega 1980.]

Wilson, N. L. (1959): "Substances Without Substrata", *Review of Metaphysics*, 12, págs. 521-39.
Wimmer, H. y Perner, J. (1983): "Beliefs About Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception", *Cognition*, 13, págs. 103-28.
Winsatt, W. (1974): "Complexity and Organization", en K. Schaffner y R. S. Cohen, comps., PSA y 1972 (Philosophy of Science Association), Dordrecht, Reidel, págs. 67-86.
Winograd, T. (1972): *Understanding Natural Language*, Nueva York, Academic Press.
Wittgenstein, L. (1958): *Philosophical Investigations*, comp. por G. E. M. Anscombe, Oxford, Blackwell.
Woodruff, G. and Premack, D. (1979): "Intentional Communication in the Chimpanzee: The Development of Deception", *Cognition*, 7, págs. 333-62.
Woods, W. A. (1975): "What's in a Link?", en B. Bobrow y A. Collins, comps, *Representation and Understanding*, Nueva York, Academic Press.
Woods, W. A. (1981): "Procedural Semantics as a Theory of Meaning", en A. K. Joshi, B. L. Webber e I. A. Sag, comps., *Elements of Discourse Understanding*, Cambridge, Cambridge University Press.
Woods, W. A. y Makhoul, J. (1974): "Mechanical Inference Problems in Continuous Speech Understanding", *Artificial Intelligence*, 5, págs. 73-91.
Zeman, J. (1963): "Information and the Brain", en N. Wiener y J. P. Schade, comps., *Nerve, Brain and Memory Models: Progress in Brain Research*, II, Nueva York, Elsevier Publishing Co. [Hay versión castellana: *Sobre modelos de los nervios, el cerebro y la memoria*, Madrid, Tecnos, 1969.]

Indice temático

ABBOTT, E.A., 287
Abeja, 217, 228-230; *véase también* Insecto
ABELSON, R., 57
Abstracción, a niveles de, 57, 63, 130, 140, 210
Abstracta, 58, 60-62, 64, 74, 91
Abstracto, 58; psicología popular, 58; la información como, 214; neurofisiología, 68; las proposiciones como, 116-119, 122; las propiedades relacionales, 82; descripción de la máquina de Turing, 71
Acceso, 181-182, 263, 271, 276, 284; privilegiado, 89, 126, 149, 190, 193, 265, 276-277, 282, 297
Acceso privilegiado, 89, 126, 149, 190, 193, 265, 276-277, 282, 297
Acción, 53, 58, 62-64, 105, 148, 265, 278
Acción automatizada, 200
Acerca de, 41, 116, 120, 125, 147, 153, 166, 170, 174, 177, 181, 182, 205; *véase también* Intencionalidad; acerquidad, 107, 214, 240, 256; la acerquidad de Pickwick, 147; la acerquidadde Papá Noel, 170, 174; acerquidad fuerte, 178; acerquidad débil, 166, 176, 180
ACKERMAN, R., 257 n.
Actitud intencional, 11, 16, 24, 25-43, 46, 55-59, 62, 72-82, 83, 91, 97, 99, 101, 103, 187, 209, 213, 219, 227, 229, 235, 274, 308
ADAMS, F., 254
Adaptacionismo, 211-212, 231-238, 245-250, 253, 275, 279 n., 282, 283
Admisión, 61, 246, 262, 265, 271, 280
Afinidad cognitiva, 181
Afirmación *a priori*, 22, 194, 202, 203, 209, 219, 233-234
Agnosia, 90
Aguafiestas, 217-221, 230
AKINS, K., 13, 23, 204, 273
ALEXANDER, P., 184 n.
Algo acerca de los pelirrojos, 138
Algoritmo, 80; nivel, 77, 79-80, 203, 206
Alma, 143, 296
Almeja, 32, 38
Alta tecnología, 76
Altruismo recíproco, 217
Alucinación, 29, 99
Ambiente: artificial, 252; relación del sistema intencional con, 40-41, 49, 55, 67-68, 71, 82, 101, 126, 131, 135, 140, 141, 143, 144, 189, 248, 267, 272, 274
AMUNDSON, R., 248
ANAXÁGORAS, 60
ANDERSON, A.R., 95
Anécdotas en etología, 220-223, 240
ANSCOMBE, G.E.M., 46 n., 299, 304
Ansiedad, los sabores de la, 219
Antropología, 53, 211, 245
Apreciación: de las razones de la Madre Naturaleza, 265, 279-280; de las razones de un organismo, 270-271;
Aprendizaje, 67, 144-145, 206, 249, 269-272
Aprendizaje del lenguaje, 77
Aproximación: de las características intencionales, 109, 187, 227 n.; *véase también* Imprecisión; de lo óptimo, 62, 66, 67, 76, 81, 84, 96, 263
Aproximaciones sucesivas, 271, 280 n.
AQUILA, R., 143, 147, 164 n., 173
Araña, 32, 64, 228 n.
ARISTÓTELES, 250, 253
Aritmética, 79, 80 n., 84-85, 99, 198-199
Armas, su carrera en la evolución, 217
Arriesgar, 163-164

Artefacto, 255-256, 260, 263, 265, 269, 272-273, 276-277, 284; hermenéutica, 269, 275-276
Artefactuales, objetos nocionales, 183-184
Artificial (*versus* natural), selección, 251, 259, 281
Astrología, 27, 53; sistema astrológico, 27
Atención, 100, 183, 217, 229; dirigiendo la atención de los teóricos, 232, 245
Atribución de la creencia, 30-35, 41, 43, 44, 48, 55, 58-64, 84-89, 92-108, 111, 123, 155 n., 162, 166, 170, 173, 204, 221, 235, 240
Atributo, 160, 164, 168-169, 176
Atril como sistema intencional, el, 33
Auto-: atribución de creencia, 22; aplicación conspicua del conocimiento, 20, 25, 52, 54, 57, 92, 95, 192, 197; control, 263; decepción, 22, 38; interpretación, 89; monitoreo, 263; referencia, 118, 127
Autómata, 17, 70-71, 76
Autofenomenología, 142
Autopsicología, 89

Bacteria, 269 n.
Balboa de 25 céntavos, 258-260, 265, 269
BALDWIN, efecto, 280 n.
Ballena, 165, 171-172
BARASH, D.P., 233
Barrera, 21
Baupläne, 249
BEATTY, J., 235
Bebé, 21
BECHTEL, W., 78 n., 82
Behavioral and Brain Sciences, 238-239, 245, 251
BELNAP, N., 95
BENACERRAF, J., 212, 216, 218, 219, 230, 239, 244, 298, 301, 305-307
BERLINER, H., 76
BERNSTEIN, L., 150
Biblioteca del Vaticano, 107
BIERI, P., 13
Biología, 15, 19, 20, 28, 52, 65, 103, 140, 141, 145, 211-220, 228n., 231, 237, 238, 245-247, 249-250, 261, 277-280, 282; *véanse también* Evolución; Selección natural
Bioquímica, 287, 290-291, 296
BIRO, J., 83 n.
BLACKBURN, S., 171
BLOCK, N., 71 n., 117 n., 300 n.
BODEN, M., 237 n.
BOER, S., 173
BOLT BERANEK y NEWMAN, 76
BOLTZMANN, máquina de, 207 n.
BORGES, J.L., 196
BOYLE, R., 23
Brainstorms, 11, 32, 45, 51 n., 73, 82, 90, 93, 102, 107, 129, 142, 152 n., 198, 206, 226, 296
BRAITENBERG, V., 268
BRENTANO, F., 27, 71, 118 n., 134, 142, 143, 147, 161 n., 240, 278, 299, 303
Bricolage, 285
Bridge, jugadores de, 196-200
BROCA, área de, 108
Broma, 78-79
BROWN, P., 254
Bugs Bunny, 186
"Burbuja", 276
BURGE, T., 12, 48, 69, 113, 119, 122, 140 n., 160, 165, 168n., 170 n., 181, 182, 255-258, 260-261, 265, 273, 274, 275, 276, 303
BURIDAN, Asno de, 208
BYRNE, R., 239

Caballo y "caballo", 260, 273
Cabeza: el significado está en la, 120; oraciones en la, 32, 47, 59, 67 n., 87, 91, 107, 108-109, 114, 117, 122, 129, 168; lista de compras en la, 281
Cableado, 200, 271
Cadena de la herencia, 136
CAIN, A.J., 236
Caja negra, 26, 47, 63, 68, 77, 230, 303
Calculadora, 75, 80, 199, 203
Cálculos, 192
Cálculos, prodigio en, 200 n.
Cambio Cambridge, 176
CAMPBELL, D., 63
Capa de transductores y efectores, 138-139, 148
Capa heurística, 187, 274
Capacidades discriminatorias, 57, 262; de las enzimas, 278, 280; del ordena-

dor de dos bits, 258
Captabilidad, 116, 119-120, 122, 122 n., 125-126, 187, 187 n., 190
Carencia de una embarcación, 169
Carácter (de un término u oración), 124-127, 131, 139, 153, 168n.
CAREY, S., 222 n.
CARGILE, J., 216, 217
Caridad, principio de, 245, 301, 305-306
CARNAP, R., 115, 123, 125
CARNE, Judy, 158
CASTAÑEDA, H.N., 123, 150 n., 170
CATILINA, 169
Causa, 51, 63, 74, 100; *versus* la razón, 85
Causal/es: explicación, 51, 62, 123; poderes del cerebro, 73, 286-288, 295; papel de las proposiciones, 117; el valor semántico como propiedad causal, 90, 132, 137, 204; significado de la teoría de la creencia, 51, 125, 154, 156-157, 181, 182, 190, 210; teoría de la referencia, 40, 154, 174, 177, 177 n., 180-181, 183
"Causalidad de la posición invertida", 295
Cerebro, 15, 16, 22, 26, 43, 54, 59, 61, 68-69, 72-74, 99, 108, 194, 205; *véase también* Sistema nervioso; la visión del ojo del, 132; ideas geniales, simulación de, 290; división, 22, 90; como un motor sintáctico, 66, 132, 133, 135, 139; lo que el ojo de la rana le dice al, 186, 187, 267; tridimensional, 288; la escritura, 47 n., 74, 124, 126, 128-129, 294
CÉSAR, Julio, 144, 189
CICERÓN, 169
Ciego/a: negación de la ceguera, 17, 22; la evolución como, 250-252, 278, 281; persona, 21, 101-102
Ciencia, 212, 237
Ciencia ficción, 17, 264
Cigarra, 236
Cinemática, 63, 108
Cognitivismo, 212, 215, 232 n.
Cognitivo: "actitud cognitiva", 237 n.; disonancia, 97; ilusiones, 17; proeza, 64, 69, 213, 228; psicología, 68, 126, 139, 160, 186, 192, 195-198, 206, 228 n., 248, 284, 292, 304n.; ciencia, 47, 90, 95, 114, 159, 187 n., 192-194, 203, 204; estilo, 129-130; psicología cognitiva subpersonal, 62, 65-69, 72, 90, 100, 209, 227 n.; sistema, 78 n., 137-138, 205, 248; ruedas, 187
COHEN, L.B., 21
COHEN, L.J., 86, 93
Colocación del reconocimiento de las enzimas, 237 n.
Combinatoria: explosión, 43; poderes de los sistemas de representación, 137
Competencia, modelo o teoría de, 63, 78-80, 149-150, 152, 196, 207, 213, 228; *véase también* Modelo de funcionamiento; microcompentencia, 228
Complejidad de los sistemas intencionales, la, 39, 43, 64, 106
Comprensión, 20, 66, 198, 213, 255
Computación, 77, 79, 205, 210; computacionalismo High Church, 205, 207 n.; el nivel de Marr de, 77, 80, 203-204; sin representación, 193
Comunicación, la, 37, 37 n., 90, 115, 212-213, 216-22, 224-227, 239, 244, 256, 263, 272
CONAN DOYLE, A., 147, 147 n.
Conciencia, la 11-12, 17, 23, 87-89, 91, 102, 183, 193, 196-197, 200, 217, 228 n., 297; falso, 53, 245
Conducta, 20, 26, 31, 46 n., 47, 51, 74, 102-105, 136, 141, 151, 181, 208, 217; lenguaje de la conducta, 213; conductismo, 15, 42, 47, 51-52, 56, 211-212, 215, 219, 222-224, 229, 231, 237, 248-250, 295, 299-300, 304; interpretación de la, 171, 217-220, 222, 246, 270; conductismo lógico, 51-52, 56; predicción de, 28, 31-39, 56, 59, 111, 121, 125, 215
Conejos, no pájaros, 25
Conexionismo (o nuevo conexionismo), 205-208, 292
Confabulación, 89
Conocimiento, sabiduría por medio del, 182
Constituido, el mundo nocional, 147, 152, 181
Contenido, 12, 49, 66 n., 67, 69, 90-91, 98, 102, 105, 107, 109, 112, 125-127,

130 n., 132, 137, 148, 168, 190, 267-268, 273-276, 285-286; *véanse también* Intencionalidad; Significado; no creído, 225
Content and Consciousness, 11, 12, 23, 32, 182 n., 209, 240, 267 n.
Contexto, 123-128, 131, 139, 158, 220
Contradicción: de la creencia, 25, 32, 38, 85, 87, 93, 100, 150-151; de la intuición, 17, 22
Contraespionaje, 187 n.
Contrafactual, 69, 97, 280
Control, 132, 204, 229, 263, 288; poderes cerebrales, 288, 294-295
Convención lingüística, 124, 127, 136
CONWAY, J.H., 44, 288
Corte Suprema, 189
Coste: y ventaja en el diseño, 56, 57 n., 95, 234, 262; de hacer negocios *versus* pérdida, 248
Creatividad, 252-265
Creencia, 20-23, 25-44, 47-68, 73-79, 83-98, 98-114, 117, 120, 125, 133-135, 146 n., 148, 151, 157-158, 160-164, 167, 202, 208-210, 215-219, 221, 229, 235, 240, 246, 260, 300, 306; caja de creencia, 101, 135-136; estados semejantes a creencias, 82; "creer*", 135; núcleo, 60-61; descontento, 208; fijación, 101, 204; general *versus* específico, 162-165, 170-171, 173-176; individuación de, 30, 48, 60, 105, 108, 114, 123-124, 128, 175-178; infinitud de, 61, 73
Creencia general *versus* específica, 162-165, 170-171, 173-176
Creencias o elementos núcleo, 60-62, 73
Creencias religiosas, 25
Criptografía, 38, 127 n.
Crusoe, Robinson, 69
Cuadro, 53, 122; *véase también* Imagen
Cualidades secundarias, 23
Cuántica, física, 19, 28, 74, 248-249
Cuantificación, 117-118, 123, 161, 164-165, 167
Cuantificando en, 171-172, 178
Cuarto chino, 285, 287
Cuarto distorsionante de Ames, 22

Chapucear, 251, 283
Charla en cadena, 206
CHARNIAK, E., 78
Chauvinismo en la atribución de la mentalidad, 71-72
CHENEY, D., 212, 239, 241, 243
CHERNIAK, C., 32, 87, 93, 95, 223, 301
Chimpancé, 17, 222, 225-226, 238, 240, 243
CHISHOLM, R., 160-162, 164, 182, 261, 298-299, 304, 305
Chivo, 189
CHOMSKY, N., 77, 196, 248, 249
CHURCH, tesis de, 71
CHURCHILL, W., 51
CHURCHLAND, P.M., 23, 31 n., 73, 104, 108, 109, 116, 118, 187, 188, 207-210, 261, 306
CHURCHLAND, P.S., 31 n., 73, 104, 109, 188, 261

Dada: en instropección, 143; como texto, 128, 131, 136, 190, 266, 281
DAHLABOM, B., 13, 249
DARMSTADTER, H., 93
DARWIN, C., 233, 250-252, 278, 283-284; "darwinismo *pop*", 272; "darwinismo común", 272
Dasein, 305
DAVIDSON, D., 12, 22, 30 n., 67 n., 103, 105, 189, 261, 298, 300-303, 305-307
DAWKINS, R., 229, 236, 239, 249-252, 264, 278, 284 n.
De re: creencia, 40 n., 105, 107; *versus de dicto*, 111-112, 152, 155-159, 160-182, 184-185, 190
DE SOUSA, R., 93
Decisión: teoría de, 63, 93-95, 230; árbol, 81
Deducción, 28, 39, 137, 161, 192; deducción, encierro, 92-93; deducción lógica, 94-95; del conocimiento, 93
Déjà vu, 231
Delfín, 103, 226, 261
DELOACHE, J., 21
DEMÓCRITO, 30 n.
Demonio, 75, 103, 154, 289
Demostrativos, los, 173
DENNETT, D., 11, 17, 23, 55, 61, 72-73, 75, 77, 82, 83, 96, 101, 103, 127, 132, 153 n., 159, 159 n., 168, 174, 186, 189, 194, 204-205, 207, 209, 212, 228, 248, 255, 263, 267, 279, 280n., 285, 300-308

Depredador, 103, 212, 217, 237, 242, 244; alarma, 217; engaño del, 229-230; detector, 201, 268, 269 n.
DERRIDA, J., 47 n., 307 n.
DESCARTES, 23, 48, 125, 148-149, 153, 193, 232 n.
Descripción, 147, 150, 173; definida, 155, 165, 173, 175, 179-181; descripción definida, teoría de la, 174-175; el conocimiento, mediante, 182
Deseo, 20, 21, 26, 30-42, 50-65, 67, 83, 86, 90-92, 101-106, 133-147, 209, 213, 216-221, 229, 233, 235, 246, 279, 300-301; estados parecidos a deseos, 82
Designación rígida, 120, 177 n.
Desinterpretación de los estados intencionales, 39, 133, 135, 267
Determinismo del mundo de la vida, 45
DEWDNEY, A., 288
DICKENS, C., 147
Dinámica, 63
Dios, 65, 142, 253, 255, 264
Directa: cita, 161, 302; referencia *versus* indirecto, 173-174; *versus* el discurso indirecto, 123, 168n., 171
Discriminación: de los significados, 65-67; umbrales, 29
Discusión mentalista, 15, 20, 23, 47, 69-70, 101, 103, 213, 215, 229, 233-235
Diseño, 42, 43, 56, 63, 228, 234-235, 250, 251, 256, 262, 264, 266, 268, 273, 281 n., 282, 292; de los sistemas de la IA, 76, 81, 95, 227-228; análisis, 69, 128, 144-145, 159, 201, 206, 207, 226; el argumento a partir del diseño, 284; bueno, mejor y óptimo, 38, 56, 234, 238, 245; de los experimentos, 209, 223-226, 239, 240, 242, 247; de los organimos por la evolución, 67, 251; autodiseño, 200; atajos, 17, 56, 227; la actitud, 28, 35, 45, 46, 75, 103, 109
Disposiciones, 47, 52, 63, 65, 107, 109, 134, 145, 148-149, 151, 190, 197
Dispositivo de dos bits, 49, 256-260, 262-265, 269, 272-277, 282-283
Distracción, 87-89
DOBZHANSKY, T., 236
Dolor, 20, 22, 59
DONNELLAN, K., 115, 154, 165, 170-171, 173
Doppelgänger, 119-120, 126-127, 139, 141, 147, 153, 165, 167
DOVER WILSON, J., 224 n.
DRETSKE, F., 48, 103, 104, 186, 214, 250, 254, 257-258, 260-261, 265-273, 275-276, 283, 303
Drosofila, 284 n.
Dualismo, 193, 229, 296
Dugongo, 172-173, 183
DUHEM, P., 58 n.
DUMMETT, M., 112, 115
DYBAS, H., 236

ECCLES, J., 261, 296
Ecología, 158 n., 162-163, 238, 268
Economía, 186, 188, 209, 246
Ecuador, el, 58, 74, 300, 305-306, 308
EDELSTEIN, D., 13
Edipo, 123
Efectores, 101, 132, 133, 139, 148, 293
Egoísta: ARN, 251 n.; gen, 251 n., 264, 279
EINSTEIN, A., 195
ELDREDGE, N., 239
Electrón, 74, 111
Embriología, 237 n.; restricciones embiológicas, 248
Emergencia, 206
Empatía, 26, 98
ENC, B., 257 n.
Engaño, 97
Engaño, el, 120, 229-230
Enigmas de los filósofos, 101-102, 109, 114, 142, 157, 183, 185, 276
Entornado, 108
Episodios mentales, 89, 183, 192, 304
Époché, o paréntesis, 142, 149
Equivocaciones, 83, 84, 86, 88-89, 98-100, 257-258, 267, 272
ERLER, A., 258
Error, 29-30, 55-57, 81, 84, 85, 89, 90, 99-100, 240, 255-256, 267, 272, 273, 276-278, 279; *véase también* Equivocaciones
Esencia, 143, 187, 282 n.; esencialismo, 175, 178
Espacio Hilbert, 74
Especificaciones, 68, 80, 146, 227
Espectador: la función a la vista del, 277-278; información en la mente del, 190;

el significado en el ojo del, 44
Espectograma, 76, 241
Esperando las proposiciones, 119, 124-125, 149
Esperanza, 20, 23, 67, 213
Esperanza por los jeroglíficos, 67 n.
Espía, 105, 107, 109, 162; el de menor estatura, 162 n., 165-166, 171, 176, 178-179, 181, 184, 185
Estadísticas, propiedades, 205-206
Estegosaurio, 283 n.
Estereo-visión, 21
Estímulo: generalización, 67; significado, 47; sustitución, 224-225
Estrangulador loco, el, 215
Estrategia Evolutiva Estable (ESS), 230, 236
Estrategia predictiva, 26-27
Estructuralismo, 248-249
Estructuras de los datos, 73
Etología, 186, 235, 238-240; cognitiva, 211, 228
EUCLIDES, 195
EVANS, G., 115, 154, 158 n., 182, 185
Evolución, 11, 18, 42, 48, 54-56, 64-66, 67, 69, 92-94, 144, 211, 217, 230-231, 232-233, 234-238, 243-246, 248-252, 264-265, 267, 272, 274, 283; *véase también* Selección natural
EWERT, J.P., 110 n.
Experiencia, 259
Experimento gavagai, 189, 243
Experimentos, 209, 223-226, 238, 239, 242, 231
Explicación: de la conducta, 56, 58; 61-63, 79, 89, 181; lo causal *versus* dar la razón, 237; histórica, 250; intencional, 68-69, 74, 158; *véase también* Niveles; mentalista, 234; psicoanalítico, 233
Explicación mentalista, 232, 235
Explícita, 60-61, 73, 137, 183, 194-196, 199-200, 204, 206; contradicción, 151-152; "explicitación", 137 n.
Expresión: de una oración de la proposición, 122, 124, 139, 154-157, 168, 188; *versus* la descripción, 137-138
Extensión, 115-116, 119, 124, 126-127, 161
Façon de parler, 33, 104; *véase también* Metafórico
Falsas: alarmas, 268; atribución de creencia, 29, 30 n., 55-57, 84-85, 125, 148, 165
Falsedad: evitando, 95; útil, 75, 94
Falso, lo, 116
Fantasma en la máquina, 193
Fatalismo, 36
FAUCONNIER, G., 191
FELDMAN, J., 254
Fenomenalismo, 153
Fenomenología (y fenomenología), 26, 79, 142-143, 146, 152, 182, 190
Fenotípica, plasticidad, 80, 281 n.
FEYERABEND, P., 27
Ficción, 44, 152 n., 186, 228; ficcionalismo, 44, 59, 75; mundo de ficción, 144, 147, 150, 183; del teórico, 112, 142-144, 278
FIELD, H., 61, 114, 116, 118, 122-123, 129 n., 130 n., 132n., 134, 136, 170 n., 189
Filética, inercia, 232
Filosofía de la mente, 22, 23, 71, 89, 298
Filosófos, 15, 16, 18, 22, 209, 254; invención de, 74; método de los, 18, 22; enigmas de, 101-102, 109, 112, 142, 157, 184, 185, 277; como sermones, 240
Finitud de los mecanismos, 80, 80 n., 138
Física, actitud, 27, 35, 44, 103, 110
Física ingenua, 21
Física, la, 15, 27-28, 44-45, 50-52, 70, 118-119, 213, 257, 259, 275; *véase también* Mecánica cuántica; dinámica y cinemática, 63; popular, 19-21, 56-62, 67-70, 74-75, 83-84, 89-92, 107-109, 203, 207-210, 247; de la rana, 104; ingenua, 21
Fisiología, 64, 69, 72, 219; *véase también* Neurofisiología
FODOR, J., 12, 18, 44 n., 48, 51, 58-59, 61, 64, 81, 91-92, 101, 104, 109, 112, 116, 119, 122-123, 126, 128-129, 132n., 142-143, 148, 184, 189-190, 193, 202, 202 n., 203-204, 207-208, 263, 265, 270-275, 283, 298, 300 n., 304-305
FOLLESDAL, D., 184 n.
Fonología, 75-76
Formal: los rasgos formalistas en filosofía, 18; coerción de la formalidad,

132 n.; formalidad de la aritmética, 79-80, 117; formalidad de los programas, 141, 204, 286, 296-297; lógica y racionalidad, 95-96
FORSTER, E.M., 107
Fotón, 103
Fotosensibilidad, 101
Fototropismo, 17
FREGE, G., 111, 115-116, 119, 121-124, 161, 165, 190
FRIEDMAN, M., 58 n., 68, 69,
Fuerza centrífuga, 23
Función, 117, 127, 130, 135-136, 248-249, 259, 269, 272-273, 274, 279, 282; *véase también* Objeto, el; asignada, 266; funcionalismo antropológico, 245; funcionalismo en la filosofía de la mente, 128, 131, 248, 255, 300, 304; caracterización funcional, 42, 70-71, 74, 128, 137, 140, 144, 147-149, 168, 201, 246 n., 253, 266, 282, 283-284; matemática, 77, 104; normal, 267; funcionalismo original, 190, 284; propia, 258, 265

GARDNER, H., 90
GARRETT, O.B., 160, 163
GAZZANIGA, M., 22, 90
GEACH, P. 176, 299, 304
Gen, 111, 252, 264, 278, 280 n.; egoísta, 251 n., 264, 278
Generatividad o composicionalidad, 43, 54, 59, 73, 138, 194, 207
Genética, 237 n., 246
Genotipo, 230, 280 n.
Gentes, 129
GHISELIN, M.T., 239, 246, 249
GIBSON, E., 21
GIBSON, J.J., 158 n., 160
GOFAI, 292
GOLDMAN, A., 73
GOODMAN, N., 147, 153, 165
GOREN, C.G., 21
GOULD, J.L., 227
GOULD, S.J., 231-237, 247, 249-250, 272, 284
GOUNOD, C., 150
Gradualismo, 248, 252
Gramática, 52, 76, 93, 194, 197, 205, 249
GRANDY, R., 302
Gravedad: centro de la, 58, 74; "La gravedad hundida", 228
Gravedad de la luz de los reflectores, 11
GREGORY, R., 22
GRICE, H.P., 65 n., 216, 221
GRIFFIN, D., 238
GROLIER, enciclopedia, 186
GUNDERSON, K., 68

Habla, acto del, 53, 173, 216, 224, 244
Hablando con uno mismo, 106, 109, 199
Hamlet, 113, 224
HAMPSHIRE, S., 37, 151 n.
Hardware, 77, 204, 206
HARMAN, G., 12, 114, 116-117, 122, 180 n., 189, 300 n., 304 n.
Harvard, 176, 231, 232
HAUGELAND, J., 184 n., 241, 254, 261, 292, 306-307
HAYES, P., 21, 40, 159 n., 184 n.
Hecho: más profundo, 38, 44, 48, 81, 110, 259-260, 266, 277, 282; histórico inerte, 144, 190, 191, 281-282; como la verdadera proposición, 114
HEIDEGGER, M., 307
Heminegación, 17
Hermenéutica, 26, 90, 128, 268; principio de la hermenéutica de los artefactos, *269*, 275-276
Heterofenomenología, 142, 146, 149, 152; *véase también* Punto de vista del tercero, el
HEYES, C., 239
Hidrocarbón gaseoso, 276
HINTIKKA, J., 147
HINTON, G., 207 n.
Hipnosis, 89
Hipótesis: confirmación, 91, 96, 218, 222-223, 236, 242-243, 246, 251 n.; generación, 79, 97, 98, 211, 218, 239, 242, 244, 247
HIRSCHMANN, D., 184 n.
Histórica; *véase también* Hecho; histórico inerte, como la explicación evolutiva, 246, 250; histórico-arquitectónico, 237
HOFSTADTER, D.R., 12, 150, 153 n., 189, 255, 261, 285, 296
Holismo, 62-63, 205, 232 n., 248
Holmes, Sherlock, 147, 222-228, 240-241, 242-243
Homúnculos, 244

HOOKE, R., 236
HOOVER, J.E., 181
HORNSBY, J., 172 n., 173
HOUSE, W., 171
HUME, D., 87
HUMPHREY, N., 217
HUNTER, I.M.L., 200 n.
HUSSERL, E., 142, 149
HWIM, 76

Idea, 53; "la mente no puede llegar más allá del círculo de sus ideas", 160, 182
Ideal: ambiente, 144, 149; la evolución como un ingeniero ideal, 283
Idealismo de la física cuántica, 19
Idealización: del nivel de computación, 204; en economía, 246; de la actitud intencional, 54-58, 62, 75, 77, 80, 90, 92, 109-110, 210, 227; problema, 305
Identidad: condiciones, 60, 161; *véase también* Individuación; equivocada, 158; personal, 153 n.; teoría de la, 69, 82 n.
Ideológico, 96-97
Idiota, 22, 225 n.
Iguana, 228
Iluminismo, 248-249
Ilusión: cognitiva 17; perceptiva, 56; filosófica, 104, 106, 120, 180, 185, 226, 246, 280
Illata, 58-60, 67, 89-91
Imagen manifiesta, 23, 209
Imagen mental, 53, 67 n., 78, 122
Imaginación, 183; en la narración de cuentos, 232, 247
Imperativo, 218, 220
Imperfección de los sistemas intencionales, 38, 56, 91, 227
Implícito, 60-61, 73, 138, 194-196
Importancia, 55
Imprecisión de la adscripción de contenidos, 104-106, 109-110, 151, 187, 274
Incompleto en matemática, 195
Inconmensurabilidad, 188
Inconsciente, 79, 89, 183, 193, 196-197, 233, 296
Indeterminación: de la función, 283; de la interpretación radical o intencional, 89, 99-100, 137, 149, 152, 189, 258, 269, 272, 274, 277, 279 n.; de la identidad personal, 47; de la física cuántica, 28; de la traducción radical, 26, 47-48, 99, 127n., 189, 212, 224, 243, 246 n., 251, 282, 299, 301 n., 302, 305-306
Indicatividad, 121, 123-124, 130, 141, 150
Individuación de las creencias, 30, 48, 60, 105, 108-109, 113, 123-124, 127, 175-178
Individualismo: el antiindividualismo de Burge, 260, 273-274; metodológico, 245
Inferencia, 83-85, 95, 137, 204
Infinitud de las creencias, 61, 73
Información, 59, 61, 95, 101-102, 105, 132, 135-141, 145, 189, 194-195, 202, 210, 243-245, 288; baño de, 101, 103, 105, 106; pendiente, 244; era de la información, 186; en la mente del observador, 190; procesamiento, 193-195, 227, 269, 289-290; las proposiciones como unidades de, 186; exigidas para predecir la conducta, 36, 37 n.; semánticas, 185-189; sensibilidad, 79, 208; almacenaje, 73, 240; teoría, 186, 214; vehículos de, 53, 67, 115, 132
Ingeniería, 35, 92-93, 98, 149, 169, 234, 251-252, 257, 262, 267, 270, 274, 280, 282-283, 291; Oscar, el ingeniero, 92-93
Iniciativa para la Defensa Estratégica (SDI), 291
Innato, 21, 271; innatismo (o nativismo), 248-249
Insecto, 32, 64, 101, 103, 205, 228 n., 230, 235
Insinceridad, 108
Instantáneas, las representaciones como, 159
Instinto, 222
Instrumentalismo, 39, 45, 58, 61 n., 62, 65, 72-75, 78, 82, 209, 274, 305
Inteligencia, 37, 37 n., 43, 51, 53, 64, 192, 217, 219, 222, 229; *véase también* Inteligencia artificial, IA
Inteligencia artificial (IA), 57, 64, 136, 147-149, 152, 159, 159 n., 168, 187, 204, 206, 227-228, 260-261, 285-288, 291-293, 295; "fuerte" *versus* "dé-

bil", 285-287, 292-293
Intención (en el significado común), 20, 22, 37, 64, 217, 220, 232, 245, 277; distinguido del sentimiento filosófico, 240; falacia intencional, 282; del autor o diseñador, 46, 257, 276, 281, 282; del usuario, 266, 269, 271-272
Intencional: idioma o lenguaje, 50, 62, 67, 78, 161, 214-216, 248-249, 277-280, 300, 304; rótulo, 67, 90, 93; objeto, 144, 146-147, 161; "segunda intención", 219; estado, 96, 216, 240, 273, 277
Intencionalidad, 11, 27, 44, 71, 72, 73, 106, 118 n., 134, 161n., 202 n., 281, 287, 294-295, 299-300, 304; derivada, 255, 257, 263, 265, 271, 274-278, 282, 296; intrínseca, 255, 258, 260, 264, 269, 272, 274-277, 282; original, 190, 255, 259-263, 265, 269, 271-275, 282, 284
Intencionalidad derivada, 255, 257, 263, 265, 271, 274-275, 277-278, 296
Intensión, 115-116, 124, 161, 166 n., 168, 174, 279
Interpretación: por los bilingües, 212; de la ficción, 147; inflacionaria, 267-268; intencional, 38, 40-41, 54, 56, 59, 62, 67, 72, 74, 78, 84, 86-87, 89-91, 99, 100, 104, 105, 119, 126, 159, 218, 226, 245-247, 260, 276, 300, 306, 308; "teoría de la interpretación", 303; interpretacionismo, 26, 34; radical, 127n., 189, 250; semántico, 26, 124, 127, 131, 134, 136, 138, 143, 145, 235
Interpretación inflacionaria, 267-268
Interrogatorio, 99, 107
Intrínseca: intencionalidad, 255, 258, 260, 264, 269, 272, 274-278, 281; propiedades, 102, 118, 259; velocidad, 293
Introspección, 23, 53, 79, 88, 107, 143, 149, 197, 200, 269, 273; *véase también* Punto de vista de la primera persona, el
Intuición, 15, 17, 19, 22, 60, 64, 96, 153, 182; 197, 259; contradictoria, 17, 22; concepto intuitivo del procedimientoefectivo, 71, 73; concepto intuitivo de la mentalidad, 73; bombeo, 17, 273, 285
Ión, 103
IRVING, J., 142
Irracionalidad, 32, 83-90, 92-97, 150-152, 248
Irreductibilidad: del punto de vista de la primera persona, 19; del nivel intencional, 71, 299, 304
ISRAEL, D., 110 n.
ITTLESON, W.H., 22

JACKENDOFF, R., 174, 190-191
JACOB, F., 283
JACOB, P., 13
JAMES, W., 142-144, 149
JAYNES, J., 12
JEFFREY, R., 63
Jirafa, 230
JOHNSON, Arte, 158
JOHNSON, L.B., 205
JOHNSON, T., 228
JOHNSON-LAIRD, P., 83
Juicio, 22, 67, 79, 96, 108, 182

KAHNEMAN, D., 57, 221 n., 247, 249
KAPLAN, D., 115, 119-122, 123-127, 129, 131, 133, 139, 153, 157, 160, 168 n., 172-174
KELLER, Helen, 142
KHALFA, J., 13
KITCHER, P., 250 n.
KITCHER, Ph, 246 n., 249
KITELY, M., 170
KLEBB, Rosa, 166, 176, 178; *véase también* Espía de menor estatura, el
KORSAKOFF, síndrome de, 90
KRIPKE, S., 23, 48, 154, 165 n., 166 n., 177 n., 250, 255, 257-258, 260-261, 276-277, 303
KUHN, T., 16, 248

LAMARCK, J., 251 n., 280 n.
LAPLACE, P., 28, 34, 35, 45, 103, 144
Laugh-in, 158
LEDOUX, J.E., 89
LEIBNIZ, G.W., 59, 238
Lenguaje: ordenador, 136; aprendizaje, 52, 77; máquina, 205-206; como modelo del sistema de representación, 43, 139, 207; natural, 129-130, 134, 137, 140, 202, 207; común, 52; papel en individualizar creencias, 31-32;

105-107, 134, 160, 183, 187n., 306; *véase también* Opiniones; seres que lo usan, 69, 91, 244
Lenguaje de los micos, 212-213, 219 n.
Lenguaje del pensamiento, 32, 43, 47, 49, 100, 106, 108-109, 111, 124, 131, 133, 136-137, 151, 168 n., 185, 194, 201, 203-210, 304-305
Leopardo, 212, 217, 224
LETTVIN, J., 110 n., 267
LEVIN, J., 302
LEWIS, D., 123, 127 n., 147, 150, 152 n., 153 n., 189, 300 n., 301
LEWONTIN, R., 231-236, 239, 247, 248-250
Ley, la: cobertura causal, 59; de la física popular, 23; en la vida, 45; de la naturaleza, 276; de la física, 27; psicofísica, 67n.
Liberalismo en la atribución de mentalidad, el, 71
Libre albedrío, 12, 19
Lingüificación, 137-138
Lingüístico: análisis, 184; convenciones, 124, 127, 127 n., 136, 220 n.; lingüistas, 249
Lisboa, terremoto de, 231
Literaria, crítica, 186, 282
LIVINGSTON, R., 102
LOAR, B., 109, 170, 189
Lobo, el chico que gritó, 222, 239
LOCKE, J., 66 n.
Lógica: autoridad de, 94; epistémica, 127; filosofía de la, 185; prelogicalidad, 301; y racionalidad, 94-97; consistencia, 92-93; forma, 174; estado, 70-71; estructura, 117-118; verdad, 187 n.
Lógico, conductismo, 51-52, 56, 63
LYCAN. W., 173, 300 n., 301 n., 303

LLINAS, R., 290
LLOYD, D., 13
LLOYD, M., 236
LLOYD MORGAN, canon de parsimonia de, 219

MACKENZIE, A., 257 n.
Macromoléculas, 278-280
MACSYMA, 80 n.
Magneto, 50, 52-53
Magoo, el miope, 150
Mal funcionamiento, 38, 95
Mamíferos, los, 171-173, 281; como sistemas intencionales, 32
Manifestación de distracción, 229
Mano oculta, 251
Mapas, las creencias como, 53
Máquina expendedora de monedas, 132, 255, 257; *véase también* Máquina de dos bits
Máquina voladora, 207
Marcación del ADN, 251 n., 253
MARCUS, R., 188 n.
MARKS, C., 22
MARLER, P., 212, 239
MARR, D., 77, 203, 274-275
Marte, 56, 64, 71 n.; intérpretes marcianos, 189; pronosticador marciano, 35-38, 43, 44, 46
Masoquismo, 22, 150
Matemáticas, fundamentos de las, 71
Materialismo, 15, 19-20, 134, 296
Materialismo eliminatorio, 53, 68, 104, 208-209, 299, 306
MAUGHAM, S., 25
MAURER, D., 21
MAYNARD SMITH, J., 230, 235, 239, 246, 249
MAYR, E., 250
MCCARTHY, J., 38 n., 127 n., 159 n.
MCCLELLAND, J., 205
MCDOWELL, J., 119, 171
MCFARLAND, D., 239
Medida: las proposiciones como unidades de medida psicológica, 116, 118, 187, 188
MEEHAN, J., 228
MEIER, A., 161-162
Memoria, 87-89, 100, 145, 183, 198, 217; deterioro de, 29; distribuida, 205; fotográfica, 22, 104; estructura de la, 95; sin comprensión, 198
Memoria distribuida, 205
Mental, el lenguaje, 59, 91, 108, 122-131, 133-135, 139, 141, 143, 151, 190, 202, 304; latín mental, griego mental, 129
Mentalismo, 233, 237-238, 278
MENZEL, E., 239
Meta, 28, 67, 144, 278; submeta, 262-263

Metáfora, 33, 220, 230, 259-260, 263, 277-278, 283; de los ordenadores, 231; la actitud intencional como, 103, 104-106
Metafísica, 116, 168, 175, 177-178, 181, 185, 194, 282
Metano, 276
Metodológico, 126, 131-133, 142-143, 146
Metodológico, solipsismo, 68, 125, 131-132, 133, 142-143, 147
Mico, 212-226, 238-245
Mieleros, 244
MILLIKAN, R., 13, 69 n., 77, 82, 110 n., 190-191, 255, 258-261, 265, 267, 277
MINSKY, M., 261
MIO, el gruñido, 241-245
Mito intelectualista, 192-194, 204, 206
Modelo: de causalidad en la conducta inteligente, 51; descrito o visible desde la actitud intencional, 35, 38, 43, 44, 46-47; reconocimiento, 67; *véase también* Competencia; desempeño en la ciencia, 89, 291-292, 295; semántico, 143, 148
Modelo de desempeño 77-80; *véase también* Modelo de competencia
"Modismo dramático", 301, 306, 308
Modo de presentación, 152
MODRAK, D., 254
Módulo, 76, 99-100, 204
Mono, 103, 218-227; mico, 212, 216, 217, 238-245
MONOD, J., 289-290
Monolingualismo, 130-131, 141
MOORE, R., 184 n.
Morfogénesis, 237 n.
MORSE, código, 128
MORTON, A., 160, 181, 183
MOSER, D., 19
Movilizador Inmóvil y Significador sin Significado, 255, 264
Multilingualismo, 130-131
Mundo bidimensional, 44, 288
Mundos posibles, 113, 117, 124, 146, 152, 176, 185, 187-188, 190; el mejor de los, 231, 234, 238, 246
Murciélago, 64, 230
Mutación, 64, 237

NAGEL, T., 18-20, 22, 48, 77, 153 n., 261
NAKHOUL, J., 76
Narcisismo, 46
National Endowment for the Humanities (NEH), 184 n.
National Science Foundation (NSF), 184 n.
Natural, clase, 120-121, 171
Naturaleza, 42, 94, 235; leyes de la, 276; Madre, 227, 230, 234, 236, 264-266, 268, 269-270, 275, 277-284
Naturalismo, 59, 63, 90, 142, 145, 149, 160, 277
NEISSER, U., 158, 160
NELSON, R.J., 160
Neuroanatomía, 15, 22, 64, 73
Neurofisiología, 15, 22, 52, 53, 74, 104, 113, 293-294
Neurona, la, 103-104; como ordenador, 289-290
NEWELL, A., 77, 227
Newfie, broma de, 78
NISBETT, R., 57, 83, 89, 93
Niveles: de explicación, 28, 44, 46, 77, 89, 203, 213; los tres niveles de Marr, 77, 274-281
Nocional: actitud, 111, 140-162, 171, 174, 176-177, 181-183, 190; *versus* relacional, 111-112, 160-162, 166-167, 169-170, 175; mundo, 111-112, 141-144, 146 n., 148-160, 167, 181, 183, 185, 189-190
Norma: psicología libre, 90; normal, 267; gente normal, 89; papel normal, 69; principio normativo, 301; sistema normativo, 58, 59; teoría normativa, 92, 93-94, 231
Notacional, variante, 128-164, 169
Nova Gene, 251 n., 253
NOZICK, R., 35, 44, 153 n., 234
Nuclear, imagen de la resonancia magnética, 99
Números, 118 n., 237

Objetividad de la atribución de creencia, 26, 35, 38-39, 44, 46
Objetivo: modelo, 43; punto de vista, 18-19
Observación de creencias, 30 n.
Observador, relatividad del, 44
Observatorio de Greenwich, 121
Ocurrencia de la creencia, 108-109

Odiseo (Ulises), 223
Oleico, ácido, 227, 229
Olvido, el, 87-90
Omniciencia, 142
Ontología, 74, 82, 111, 161, 164, 181, 190, 208, 301, 306-307
Opacidad, 123, 160-161, 163, 169, 181, 214, 215; 299; *versus* transparencia, 173-174, 181
Operacionalismo, 73
Opinión, 31 n., 106, 183, 188 n., 208, 299
Optimalidad, 76, 81-82, 84, 144, 231-235, 237, 245-248, 272-275; suboptimalidad, 84, 96-97
Optimización, 59, 235, 246
Oración, 123; eterna, 123-124; en la cabeza, 32, 100, 106, 108-109, 113, 117; las proposiciones como, 168, 184; comprensión de la, 66
Oraciones eternas, 123
Oracionismo, 91, 129, 137-138, 195, 204; definido, 128; actitudes oracionales, 111, 122-140, 142, 147
Orangután, 243
Orden, la, 132, 133, 148, 217, 220, 241, 244
Ordenador, 38, 43, 45, 72, 75, 81, 148, 193, 202, 205, 219, 232n., 255, 269, 277, 285-294; *véase también* Ordenador jugador de ajedrez; análogo, 97; óptico/a, 289; orgánico, 289-291; predicción de, 28, 34; programa, 45, 70, 287-297; lenguaje de programación, 136, 202 n.; "el programa correcto", 287, 290-291, 293; superordenador, 290
Ordenador jugador de ajedrez, 76, 81, 88, 256
Ordenador macromolecular, 289-291
Orgánica, contribución, 126, 131, 141, 143-146
Original: funcionalidad, 190, 284; intencionalidad, 190, 255-263, 265, 271-275, 281, 284
Ortostática, hipotensión, 288
OSTER, G., 235
Oxford, 32, 37 n., 115, 158 n.
Paisaje adaptativo, 280 n.
Pájaro, 32, 64, 103, 229-230, 233, 247, 252, 270-271; avoceta, 224n.; cuclillo, 270-271, 277; pato, 66-67n.; águila, 212; paloma, 17; chorlito silbador, 239; zancudo, 224 n.
Panamá, 258-260, 265, 274
Panda, el pulgar del, 282-283
PANGLOSS, 231-238, 245-248; "perspectiva panglossiana modificada", 208
Papá Noel, 59, 170-171, 174, 189
Paralelogramo de fuerzas, 63, 74
Paramecánicas, hipótesis, 193
Paramecia, 104, 106
Paranoia, 141, 150
PARFIT, D., 47-48, 153 n.
PARKINSON, enfermedad de, 288
Pasar por alto, 88, 99
Pata: cuatro patas, 237- 247; como categoría funcional, 246; corta, 251
Patología: cognitiva, 17, 38, 89, 97; inducción, 57, 159
PEACOCKE, C., 184 n.
Pensamiento, 87-89, 91, 125-127, 182-183
Pensamiento, experimento del, 17, 44, 49, 56, 127, 127 n., 135, 139, 141, 145, 154, 212, 251, 264, 275, 286
Pensamiento (fregeano), 111, 115, 120-121, 124
Pensar, 91, 106, 124-125, 182, 192, 204; *versus* creer, 109, 182-183
Peñique, Amy el, 178-179
Percaballo, 261
Percepción, 52, 55-56, 64, 86, 102, 148
Percepción aparente, 21
Periferialismo, 47
Periféricos, orgános y módulos, 101, 199, 204
PERNER, J., 21, 239
Perro, 67 n., 129, 183, 208, 216-217, 244, 251
Perruno, 129
PERRY, J., 115, 119, 121, 123, 129, 133, 150, 153 n.
Perspectivalismo, 47, 68
Pescado, 32, 237, 261, 268
PHILBY, Kim, 186
PICKWICK, acerquidad de, 147, 171, 174
Planeamiento, 292
Planeta Tierra, Gemelo, 41, 49, 119-123, 127, 142, 143, 146, 147, 153, 165, 176, 185, 189, 258, 260, 275
Plantas como sistemas intencionales, las, 33

Plasticidad, 145, 246
PLATÓN, 50, 116
Plaza Trafalgar, 60, 188
Plurales, lógica de los, 172 n.
Pluralismo, 231, 233
Poirot, Hercule, 180, 181
Polo Este, 299 n.
Polo Norte del mentalismo, 299, 303, 305
POPPER, K., 261
Popular: física, 20-24, 50, 54; psicología, 19, 21, 56-62, 64, 67-70, 74-75, 83-84, 88-92, 107-109, 203, 207-210, 247
¿Por qué?", preguntas, 30, 237-238, 246, 249-250, 253, 278
Positivismo, 16, 73, 232, 237; pospositivismo conservador de Harvard, 232
Positivismo lógico, 16
POWERS, L., 93
Pragmática, implicancia, 179, 302
Pragmatismo, 260, 302
Precisión: de los estados cerebrales, 61-62, 91; mal colocada, 60; de la proposición, 117
Predicado, cálculo del, 186-187
Predicción: desde la actitud de diseño, 104, 209; desde la actitud intencional, 26-27, 28-29, 30-39, 42, 46, 54, 59, 61, 63, 71, 76, 79-82, 83-84, 104, 120, 157-158, 181, 213, 219, 228, 235, 247, 280; del comportamiento de la máquina de Turing, 70-71; del tiempo, 290
Pregunta: conceptual, 50-51; empírica, 219, 225 n., 229; vacía, 48; ontológica, 74; abierta, 95; socrática, 50-51; la buena de Stich, 99; "¿por qué?", 30, 237-238, 246, 249-250, 253, 278
PREMACK, D., 222, 225, 226 n., 238
Presión del tiempo, 81, 95; *véase también* Velocidad
Presión selectiva, 237, 246
Primate, 239; *véanse también* Chimpancé; Mono; Orangután; Mico
Principio de humanidad, 302
Principio factunorma, 96
Principio proyectivo, 302, 306
PRIOR, A., 165
Privacidad, 192
Probabilidad, 59, 63, 92, 138, 252
Problema de la disyunción, 257, 265-277, 278
Problema del armazón, 204
Problemas, resolución de, 61, 67, 183
Procedimiento: bueno, 86; verificación, 115; semántica del procedimiento, 304
Procedimiento efectivo, 71, 73
Procesamiento paralelo, 288-292
Proceso, estar al día como, 158 n., 159, 183
Proposiciones, 113, 184-190, 276; como dólares, no números, 188; no oraciones, 138-139; actitud proposicional, 38, 60, 78-79, 93, 100, 103, 105-112, 114, 204, 208; cálculo proposicional, 186
Propósito, 256, 258-259, 264-266, 268, 271, 281, 283; *véanse también* Meta; Función; Intención; *Raison d'être*; Teleología; propositación, 281
Prostética, visión, 101-102
Proyectabilidad, 64, 66, 118-119, 133, 181, 190
Psicoanálisis, 233, 284 n.
Psicología, 51-52, 68-69, 87, 89, 116, 117, 119-120, 125, 133, 139, 148, 153, 157-160, 183, 184-185, 239, 247, 274; académica, 22, 91, 110 n.; animal, 101-104, 238; auto—, 89; popular, 19, 21, 56-62, 64, 67-70, 73-74, 83-84, 89-92, 107-109, 203, 207-210, 246; estrecha, 126, 131, 139-141, 143, 146, 152, 160, 170, 175-177, 185, 189, 208, 245; naturalista, 142-143, 160; actitud nocional, 147, 181, 189-190; filosofía de, 185, 191; fiolosófica, 26; actitud proposicional, 74, 106, 113-114, 187, 204; racional, 142-143; semántica, 133, 137; actitud oracional, 128, 142, 147, 156, 160; social, 186; subpersonal cognitiva, 62, 65, 66, 67, 69, 72
Psicológica, semejanza, 150
Psicológico, estado, 119 n., 119-121, 123-126, 129-130, 140, 157, 164 n., 166, 168, 174-175, 181, 187 n., 188
PUCCETTI, R., 296
Punto de vista: de la primera persona, 102, 296; en el mundo de la vida, 46; marciano, 35; objetivo, 18-19; en la visión prostética, 102; subjetivo, 19, 124; del tercero, 20, 102, 142, 146, 294

PUTNAM, H., 12, 30 n., 41, 49, 52, 68, 111, 112, 119-121, 126-127, 128, 133, 139, 141, 144, 146, 147, 153, 153 n., 163, 171, 254, 275, 276, 298, 300, 300 n., 303, 305, 306
PYLYSHYN, Z., 78, 160, 184 n.

Qualia, 12, 102, 133, 168 n., 174, 209
Química, 19, 50-52, 56, 65, 150, 213; *véase también* Bioquímica
QUINE, W.V.O., 23, 30 n., 47, 99, 102, 105, 107, 112, 122-124, 160-165, 168-169, 172 n., 181, 189, 212, 220 n., 224-225, 246 n., 261, 276, 282 n., 298-308

Racional, agente, 27, 28, 56, 209, 240, 255,
Racionalidad, 43, 57-58, 81-82, 90, 92, 94, 142-143; presunción de, 27, 28, 32, 33, 37, 55-56, 72, 94, 97, 209, 216, 220-221, 226, 229, 231, 235, 238, 245-246, 275, 301-302; ¿definida cómo?, 92-97; imperfecta, 226-227; perfecta, 38, 57
Racionalismo, significado del, 190, 277, 282 n.
RACHLIN, H., 239
Radical, traducción, 26, 47-48, 99, 128 n., 189, 212, 224, 243, 246, 246 n., 251, 282, 299, 301 n., 301-303, 305-306
Raison d'être, 258, 262, 264-265, 282
RAMACHANDRAN, V.S., 78, 275
Rana, 98, 101-110, 186, 187, 208, 267
RAPHAEL, B., 76
Rayo como sistema intencional, el, 33
Razón: para creer en la creencia, 111; para impartir o retener información, 244; para cometer un error, 85; no representada por la evolución, 251, 264-265, 275, 281-282; práctica, 64; real, 233, 237; "ellos no tienen razón por la cual", 279; *versus* causa, 85
Razón de ser, 230, 237-238, 267; de flotación libre, 230, 250-251, 253, 271
REAGAN, R., 291
Real: significado, 258, 260, 271, 276, 284; *véase también* Significado intrínseco; "patentemente real", 18; *realmente* cree, 88, 90, 98, 100, 102, 104, 110, 303; *versus* aparente, 133; *versus* "como si", 263-265, 278
Realismo, 26, 44, 46, 47, 58, 73, 75, 81, 99, 152
Realismo (como diferente del realismo), 74, 82, 98, 105-106, 108-109, 203-204, 207, 274-276, 284, 305-306; "antirrealismo", 74
Realización: de los estados de creencia, 123, 128; de un sistema intencional, 71, 90, 229; de los estados *Q* y *QB*, 258; de una máquina de Turing, 70-71
Reculer pour mieux sauter, 251
Recursión, 217
Reduccionismo, 51-52, 53, 62, 69-72, 73, 74, 300
Referencia, 41, 122-123, 133-134, 153, 158, 165-166, 169, 173, 175-176, 180-181, 183, 189, 201, 214-215; teoría causal de la, 40, 154, 174, 177, 177n., 181-182, 183; directa *versus* indirecta, 173-174, 177, 178-179, 190; pura *versus* impura, 169-170; especial, 175, 176-178, 180
Refuerzo, 42, 225, 233, 237, 246
Registro, 305
Regla, 192, 198-201, 206, 260
REICHENBACH, H., 58
Relación, 164; en intensidad, 168; relacional *versus* nocional, 112, 160-162, 166-172, 174, 175, 178, 181; parecido a una relación, 147
Relativismo, 26, 27, 34-35, 38
REP, 135, 137
Representación, 41, 43, 64, 74, 91, 100, 108, 114, 118 n., 122, 128, 132, 138, 149, 152, 157-158, 183, 192, 202, 206, 210, 217, 226, 228, 269, 303, 306; falsa, 255, 257-258, 266-272, 276; la no representación en la evolución, 251, 265, 275, 280-281; teoría representacional de la mente, 272
Representados, los, 143, 150
Reptil, 32
Respuesta, 41, 222
RISTAU, C., 239, 254
Robot, 41, 43, 75, 86, 148-149, 261-264, 267, 270, 280-281; robots, teatro de, 158; mico, 243
ROITBLATT, H., 211

Romanticismo, 249
Romántico (*versus* aguafiestas), 217-220, 227
RORTY, R., 46, 254, 261
ROSENBERG, A., 107, 232 n., 234 n., 250, 278-280
ROSS, L.D., 57, 83
Ruido, 137
RUMELHART, D., 205, 207
RUSSELL, B., 113-114, 156, 173-175, 185, 215; principio de, 182, 185, 190
RYLE, G., 51-53, 63, 88, 192-198, 199, 202, 206, 295-296, 306-307

Sabiduría, 20, 22, 29, 98; por conocimiento, 182; contrastado con la creencia, 120; de la psicología popular, 52; de la gramática, 52, 76; nivel del conocimiento, 227; de ignorancia, 243-244
Sacar conclusiones precipitadas, 57, 94
SACKS, O., 22
Sadismo, 150
Saltación, 248, 252
Sapo, 110 n., *véase también* Rana
Satisfaciente, 57, 59, 178, 234, 282
SAVAGE-RUMBAUGH, S., 226 n.
SAYRE, K., 186
SCHANK, R., 57, 228
SCHEFFLER, I., 58 n.
SCHIFFER, S., 122, 152, 179, 182
SCHMENGLISH, 127
SCHULL, J., 229, 281 n.
SEARLE, J., 48, 72, 73, 77, 152, 166, 173, 180 n., 240, 250, 255, 257-261, 265, 271, 276, 285-286, 292-296
Secretaría de Estado (como agente de la KGB), 33-35
SEJNOWSKI, T., 206, 289
Selección natural, 67, 94, 230-231, 236, 250-252, 259, 264-265, 277, 279-281; *véase también* Evolución
SELLARS, W., 23, 112, 130, 167 n., 261, 298, 300-302, 305, 307
Semántica, la, 39, 40, 66, 69, 72-73, 111, 130-132, 134, 138-141, 147-148, 158-159, 161, 168 n., 194, 205-206, 213, 228, 271-272, 285, 295-296, 305; papel conceptual limitado, 189; de procedimiento, 304
Semántico/a: motor, 65, 67, 72, 132, 227 n., 296; información, 186, 188; interpretación, 68, 125, 143; red, 136; psicología, 133; valor, 132; "semanticidad", 296, 305
Sentido: datos, 133; encontrando sentido a la conducta, 54, 84-98; órgano, 100-103, 144, 204; en el sentido de Perry, 124, 127-128
Sentido común, el, 17, 74, 209
Señal, 131-132, 134-135, 266-267
Señalado, 40, 82, 88
"Servir", 90, 139
Seudocreencia y seudoconocimiento, 93
SEYFARTH, R., 212, 215, 220, 239, 241, 242-243
SHAFTZ, M., 21
SHAHAN, J., 83
SHAKESPEARE, W., 150, 224 n.
Shakey (el robot), 75
Shakey, la pizzería de, 154-157, 159, 171, 180, 181
SHANNON, C., 213
Sherlock, 60, 62, 70, 188; *véase también* Holmes
SHOEMAKER, S., 30 n.
SHRDLU, 148-149
Significado, 40, 44, 47, 64, 65, 133, 134, 210, 258, 267-271, 277, 282; para un organismo, 270; para un sistema, 269, 270-271; funcional definido, 266; de la vida, 19; racionalismo del significado, 190, 277, 282 n.; natural y no-natural definidos, 65n.
Símbolo, 43, 45, 69, 114, 122-123, 128, 130, 134, 135, 160, 194-195, 269, 287; manipulando sistemas, 203, 205-206; *véase también* Tipo
Simetría, detector, 268-269
SIMMONS, K.E.L., 229
SIMON, H., 57
Simulación, 97-98; de un creyente por otro, 98; del cerebro en el tiempo real, 289-290
Sinn, (pecado) original, 177 n.
Sintaxis, 126-128, 130-133, 135-137, 140, 142, 143, 168, 285-286, 296, 304; motor sintáctico, 65-66, 72, 111, 131-133, 227n., 271; estructuras del cerebro, 118, 140, 148, 194, 202, 228n.; teoría de la creencia, 208
Sistema intencional, 11, 16, 27, 29, 33,

35, 38-42, 44, 55, 57, 62-65, 71, 77, 90, 106, 211-238; caracterizable, 71, 73, 77, 83, 91; de un orden más alto, 216-227, 230, 239-241, 244-245; imperfección de, 39, 56, 90, 228
Sistema nervioso, 23, 103, 131-133, 136, 139, 228; *véasetambién* Cerebro
SKINNER, B.F., 42, 209, 232-235, 237, 239, 246, 248-249, 280, 295
Sloan Foundation, 184 n.
SMITH, S.B., 22
SMOLENSKY, P., 207
SOBER, E., 56, 69, 250 n., 280, 283 n.
Sociobiología, 217, 232 n., 233 n.
SÓCRATES, 50-51
Soliloquio, 229
Solipsismo, 68
Solución temporaria en la evolución, la, 55, 234
SORDAHL, T.A., 224 n.
Sordo, punto o abertura, 99, 226-227
Sorpresa, 84
SOSA, E., 167, 173, 180-181
SPENCER, H., 42, 144
STABLER, E., 194
STACK, M., 31 n.
STALNAKER, R., 114, 147, 187 n., 189, 191, 261
STANTON, Harry y Betty, 13
Star Wars, 295
STICH, S., 12, 23, 67, 83-87, 90-98, 98-99, 101, 102, 104, 109, 115, 119, 146, 184, 207-208, 254, 298, 300-302, 305-306
STRAUSS, M.S., 21
STRYER, L., 278
Subconsciente, 91
Subdoxástico, estado, 67, 93
Subjetividad, 19, 26, 38, 46, 124
Subpersonal, psicología cognitiva, 62, 65, 66, 67, 69, 72, 90, 100, 209-210, 227 n.
Superveniencia, 139
Supervivencia, 59, 64, 69, 144, 233-234; máquina de, 264, 267
Sustitución *salva veritate*; 173-174, 214-215, 278-279; *véase también* Opacidad

Tabla de escritura espiritista, 37 n.
Tácita, 57, 100, 183, 196, 198, 200-201
Tácticas, consideraciones en filosofía, 17, 18, 20
TALES, 53
TARSKI, A., 123, 134, 143, 149
Tautología, 92, 94, 183, 233-234
Taxonómico, enrejado, 136
TAYLOR, C., 137 n., 300, 304, 305
Tejido maravilla, 207
Teleología, 246, 249, 272-273, 283, 304-305
Teleportador, 153 n.
Temor, 20, 242
Tendencia genética, 232
Teneduría de libros cognitiva, 93, 248
Tener importancia, 191, 269
Teoría, 16; de la creencia, 112; ¿es el sistema una teoría?, 235; creencias cargadas de teoría, 30 n.
Teoría de la función recursiva, 71-72
Teoría del juego, 62, 92-93, 230, 233n.
Tercera persona, punto de vista de la, 16-17, 18, 20, 102, 142, 146
Termostato, 34, 35, 38-43
Teóricas, entidades, 60, 122, 304
Texto, 128, 131-132, 136, 146, 147 n., 190, 266, 281-282
THOMASON, R., 87, 172 n.
Tierra, la, 16, 56, 139, 142, 147, 154, 189, 261; terráqueo, 36-38, 127
Tipo, *véase también* Símbolo; reparto de papeles; 129-131, 140-141, 194; tipo de carácter *versus* tipo sintáctico, 126-128; tipo de hecho, 267; de oración, 123-124; teoría identificatoria del tipo, 69
Tolomeicos, epicistas, 109
TOLSTOY, L. 113
TOURETZKY, D., 207 n.
Traducción radical, 26, 47-48, 99, 128 n., 189, 212, 224, 243, 245, 246 n., 251, 282, 299, 301 n., 301-303, 305-306
Transductor, 131, 133, 137, 139, 148, 201, 257, 268-269, 293
Transitoria, representación tácita, 200
Trepando hacia arriba, 207
TRIVERS, R.L., 217
Troceado, 217
Tropismo, 219, 222, 230
TURING, máquina de, 45-46, 70-71, 73, 75, 141, 288

TVERSKY, A., 57, 123, 247, 249

ULLIAN, J., 165
Umwelt, 244
Unicornio, 165
Unificación de la ciencia, 299, 307

Valencia, 50-52, 56, 65
Vax 11/780, 75
Velocidad, en relación con la inteligencia, 287-289, 293, 296
Vendedor de limonada, 84-87, 89, 90-91, 99
VENDLER, Z., 177, 182
Venera, de ojos azules, 101
Ver, 101; *véase también* Visión
Verdad, 30, 75, 125-126, 147; atribución de la creencia verdadera, 30, 56-57, 94-95; teoría de la coherencia, 153; condiciones, 40; correspondencia, teoría de la, 115; necesaria, 43; valor, 115, 121n., 123-124; creer con reserva, 75
Verdadero, lo, 116
Verificación, 115, 165
Verificacionismo, *153*
Veritas cum grano salis, 75
Verstehen, 26, 245
Vida, el juego de la, 44-46
Videodisco, 186
Virtuales: creencias, 61, 62, 73, 186; máquinas, 206
Visión, 101-102, 139, 275
Visual, abismo, 21
Víbora, 162-163, 172 n., 212
Vívido, nombre, 152, 172-174
Voluntad, debilidad de la, 22
VON KEMPELEN, W., 76
VON NEUMANN, J., 288

WALLACE, J., 164
WASON, P., 83
WATSON, J.D., 295
WEAVER, W., 213
WEISKRANTZ, L., 22
WEIZENFLED, J., 84
WERNICKE, área de, 108
WERTHEIMER, R., 96
WHEELER, S., 47 n., 99, 307 n.
WHITEN, A., 239
WILSON, E.O., 227, 232 n., 235
WILSON, N.L., 301
WILSON, T., 89
WIMMER, H., 21, 239
WIMSATT, W., 140, 274
WINOGRAD, T., 148, 149
WITTGENSTEIN, L., 276, 304
WOODFIELD, A., 13, 111 n., 184, 255
WOODRUFF, 222, 225, 238
WOODS, W., 76, 136

YOCASTA, 123

ZEMAN, J., 129
Zombi, 296