

Cuadernos Teorema

27

INTELIGENCIA ARTIFICIAL Y EL CONCEPTO DE MENTE

A. Newell

Cuadernos Teorema

INTELIGENCIA
ARTIFICIAL
y
EL CONCEPTO
DE MENTE

A. Newell

Serie de Psicología

Director: Julio Seoane .

**INTELIGENCIA ARTIFICIAL
Y
EL CONCEPTO DE MENTE**

Cuadernos Teorema

**VOLÚMENES APARECIDOS SOBRE
INFORMÁTICA E INTELIGENCIA ARTIFICIAL**

1/A. M. Turing

¿PUEDE PENSAR UNA MÁQUINA?

8/K. Gödel

SOBRE PROPOSICIONES FORMALMENTE INDECIDIBLES
DE LOS PRINCIPIA MATHEMATICA Y SISTEMAS AFINES

27/A. Newell

INTELIGENCIA ARTIFICIAL Y EL CONCEPTO DE MENTE

VOLÚMENES EN PREPARACIÓN

A. Newell, J. C. Shaw y H. A. Simon

EXPLORACIONES EMPÍRICAS CON LA MÁQUINA DE
TEORÍA LÓGICA. UN ANÁLISIS DE HEURÍSTICA

H. Wang

HACIA UNA MATEMÁTICA MECÁNICA

E. A. Feigenbaum

INTELIGENCIA ARTIFICIAL: TEMAS EN LA SEGUNDA DÉCADA

J. A. Robinson

UNA LÓGICA DE ORIENTACIÓN MECÁNICA
BASADA EN EL PRINCIPIO DE RESOLUCIÓN

S. Watanabe, D. Mac Kay, G. A. Miller

POSIBILIDADES Y LÍMITES DE LA INTELIGENCIA ARTIFICIAL

ALLEN NEWELL

Inteligencia artificial y el concepto de mente

Introducción de

JULIO SEOANE

Traducción de

Julio Seoane y Elena Ibáñez

REVISTA TEOREMA

VALENCIA

1980

Este ensayo de A. Newell apareció originalmente publicado con el título "Artificial Intelligence and the Concept of Mind" en el libro compilado por Roger C. Schank y Kenneth Mark Colby, *Computer Models of Thought and Language*, San Francisco: W.H. Freeman & Co., 1973, pp. 1-60. La casa Freeman ha concedido a *Teorema* los derechos de la traducción castellana.

DEPARTAMENTO DE LÓGICA
DE LA
UNIVERSIDAD DE VALENCIA

PRINTED IN SPAIN

IMPRESO EN ESPAÑA

I.S.B.N. 84-370-0127-7

DEPÓSITO LEGAL: V. 123 - 1980

ARTES GRÁFICAS SOLER, S. A. - OLIVERETA, 28 - VALENCIA (18) - 1980

INTRODUCCION

El trabajo de A. Newell que aparece en este Cuaderno Teorema sobre la problemática de la Inteligencia Artificial, constituye un claro intento de organización conceptual de un área de estudio que ya cuenta con realizaciones importantes. Sin embargo, la escasa difusión que tienen estas realizaciones en la literatura científica de lengua castellana, transforma la conceptualización de Newell en una introducción abstracta a la Inteligencia Artificial dentro de nuestra comunidad. Quizá sea útil, en consecuencia, aportar algunos comentarios previos al trabajo para matizar su aparentemente nuevo perfil.

En primer lugar, es importante tomar nota de que las aportaciones de Newell a la Inteligencia Artificial y a la psicología teórica que la representa cuentan aproximadamente con unos veinticinco años. Ya hacia 1954, estaba impresionado por la noción de computador como procesador no numérico de información; esta concepción es la que explica su interés por la programación digital de computadores para jugar al ajedrez, mientras investigaba en la RAND Corporation. En esta misma época, la psicología conductista estaba en su auge de aceptación, al mismo tiempo que alcanzaba su techo teórico como explicación general de la conducta; aceptación científica y techo teórico que produjeron un efecto paradójico en los interesados por las llamadas "ciencias del computador": mientras unos seleccionaban tareas típicas del conductismo como meta de simulación (por ejemplo, E.A. Feigenbaum en 1956 con su programa EPAM para simular aprendizaje verbal (Feigenbaum, 1961)), otros como Newell, a través de la programación del juego de ajedrez y otras tareas, se planteaban el estudio de los procesos y estados no directamente observables

y se introducían en el campo tanto tiempo olvidado de los procesos de pensamiento.

Toda la investigación de Newell se puede interpretar bajo estas dos perspectivas metodológicas: una fuerte reacción contra la rigidez asociacionista y una búsqueda de las bases de lo que ahora se denomina “procesamiento de información en psicología”. Según sus propias palabras, “el conductismo y la aceptación de las normas de las ciencias naturales en psicología limitó mucho, durante una generación o más, el rango de los fenómenos de la conducta que el psicólogo, como científico, estaba deseando atacar. A menos que un aspecto de la conducta se pudiese examinar en el laboratorio y se pudiese registrar y medir de una manera objetiva, no era, según el punto de vista predominante, una materia apropiada de estudio” (A.Newell y H.A.Simon, 1961). En “Elements of a Theory of Human Problem Solving” (Newell, Shaw y Simon, 1958), se presenta ya un proyecto de investigación que rebasa con mucho cualquier marco conductista, en la medida en que se pretende una explicación de la conducta observada mediante la programación de los procesos primitivos de información que generan esa conducta.

A partir de 1955, Newell se asocia en casi todos sus trabajos con otros dos investigadores: J.C.Shaw y H.A.Simon. En cuanto al primero, estaba interesado en los sistemas de programación del computador construido a principios de los años cincuenta en la RAND Corporation, el llamado JOHNNIAC. Herbert A.Simon había comenzado a publicar más de una década antes que Newell, interesado en la toma de decisiones y en la conducta administrativa (H.A.Simon, 1944; 1947). Según comenta el mismo Simon (1972), al principio sintió cierta afinidad entre sus intereses y los trabajos de Carnap en lógica, Rashevsky en biofísica matemática y Schultz en economía matemática y estadística, e indirectamente conoció a Lotka y a Gödel.

Las contribuciones más conocidas de este equipo, cuyo análisis excede la finalidad de estos comentarios, fueron el programa LT ó *Teórico Lógico (Logic Theorist)* cuyo propósito consis-

tía en la prueba de teoremas del cálculo proposicional; el desarrollo de lenguajes de programación no numérica sino de procesamiento de listas, como el IPL (*Information Processing Language*), cuya estructura se adapta a la simulación de modelos de comportamiento; y, más recientemente, la formulación de una teoría y su correspondiente programación sobre solución de problemas, el GPS o *Solucionador General de Problemas (General Problem Solver)*.

La concepción de Inteligencia Artificial que Newell desarrolla en el trabajo que aparece en este Cuaderno Teorema, se podría quizá resumir diciendo que es un lenguaje psicológico que considera al hombre como un procesador de información. Sin embargo, ni la evolución histórica ni la pluralidad de campos que se aglutinan bajo el rótulo de Inteligencia Artificial, permiten un reduccionismo psicológico tan radical. En consecuencia, Newell dibuja tres ejes o perspectivas donde pueden encuadrarse todas las investigaciones sobre Inteligencia Artificial:

- a) en primer lugar, el estudio de los mecanismos que pueden realizar determinadas funciones intelectuales, entendiendo por mecanismo cualquier proceso abstracto que pueda realizarse, en principio, mediante un proceso físico (por ejemplo, juego de ajedrez, prueba de teoremas, etc.);
- b) en segundo lugar, Inteligencia Artificial puede entenderse como la investigación sobre métodos, recetas o procedimientos para realizar cierta clase de tareas, siempre y cuando esos métodos necesiten muy poca información para obtener resultados (métodos débiles);
- c) en tercer lugar, Inteligencia Artificial se puede entender como un tipo de psicología teórica, cuyo presupuesto fundamental es concebir al hombre como un sistema de procesamiento de información y, por tanto, que intenta estudiar el fundamento de la mente por medio de mecanismos efectivos.

Sin embargo, como ya señalamos en otro lugar (Seoane, 1979), las tres coordenadas de Newell se fundamentan en un

mismo supuesto, que también es común a la orientación de procesamiento de información; es decir, el supuesto de las etapas, que mantiene que el tiempo transcurrido entre el estímulo y la respuesta (entre el "input" y el "output", si se prefiere) se puede dividir en intervalos más pequeños, correspondiendo cada uno a algún subconjunto de acontecimientos que intervienen entre el estímulo y la respuesta; una etapa o nivel de procesamiento se corresponde generalmente con alguna representación de la información del estímulo.

Pero bajo el mismo supuesto se pueden desarrollar dos líneas distintas de investigación, cuya convergencia es deseable: por un lado, el trabajo de la Inteligencia Artificial, construyendo modelos que se realizan en programas concretos de computador, y cuya utilidad se verifica en su misma eficacia o en la comparación con fragmentos de conducta humana; por otro lado, la psicología cognitiva actual, que partiendo de modelos teóricos y de experimentación clásica, pretende alcanzar una estructura científica de los fenómenos psicológicos.

Por último, es conveniente destacar que la parte final del escrito de Newell está dedicada a uno de los problemas teóricos de la psicología, que después de muchos años de abandono vuelve a surgir con una gran complejidad de tratamientos (Pinillos, 1978): el concepto de mente. La formulación de Newell está lo suficientemente bien planteada, como para llegar a la conclusión de que "la Inteligencia Artificial o bien es irrelevante para los problemas planteados o plantea respuestas en forma inaceptable". Es decir, el concepto de mente dentro de la Inteligencia Artificial tiene un tratamiento tan radicalmente distinto al enfoque tradicional que resulta difícil cualquier traducción. Este nuevo enfoque se fundamenta en tres puntos de partida:

1. La mente es un sistema de procesamiento de información.
2. Las mentes están realizadas en sistemas físicos.
3. La neurofisiología es casi totalmente irrelevante para la naturaleza de la mente, de la misma forma que la electrónica es casi totalmente irrelevante para la naturaleza de la

Inteligencia Artificial.

Bajo tales supuestos, resulta difícil determinar la posición de esta perspectiva dentro del marco clásico de reduccionismo-emergentismo. Pero la consecuencia más notable, sin embargo, es que los viejos conceptos desterrados tradicionalmente de la psicología vuelven a surgir ahora, aunque con una elaboración radicalmente distinta.

J. SEOANE

REFERENCIAS

- FEIGENBAUM, E.A.(1961): "The Simulation of Verbal Learning Behaviour", *Proceedings of the Western Joint Computer Conference*, 121-132.
- NEWELL, A., SHAW, J.C., SIMON, H.A. (1958): "Elements of a Theory of Human Problem Solving", *Psychological Review*, 65, 151-166.
- NEWELL, A., y SIMON, H.A. (1961): "Simulación del pensamiento humano", (1961). Revista *Teorema*, volumen IV (año 1974), número 3, páginas 335-378. Traducción castellana de Julio Seoane Rey.
- PINILLOS, J.L. (1978): *Lo físico y lo mental*. Madrid, Boletín Informativo Juan March, n.71.
- SEOANE, J. (1979): *Inteligencia Artificial y Procesamiento de Información*. Madrid: Boletín Informativo Juan March (en prensa).
- SIMON, H.A. (1944): "Decision Making and Administrative Organization", *Public Administration Review*, 4, 16-31.
- SIMON, H.A. (1947): *Administrative Behavior*, New York: MacMillan.

**INTELIGENCIA ARTIFICIAL
Y
EL CONCEPTO DE MENTE**

Allen Newell

Carnegie-Mellon University

La naturaleza de la inteligencia artificial

Primera concepción: La exploración de funciones intelectuales

Segunda concepción: La ciencia de los métodos débiles

Generar y Comprobar
Hipótesis y Comparación
Métodos débiles

Tercera concepción: Psicología teórica

Extrapolación de secuencia
Solución de problemas en criptoaritmética
Memoria inmediata
Psicolingüística
Aprendizaje verbal y percepción
Inteligencia artificial como psicología teórica
Sumario

Cuestiones sobre la naturaleza de la mente

Cuestiones sobre la naturaleza general de la mente

Argumentos analógicos

Cuestiones sobre la mente de los computadores

Cuestiones sobre la naturaleza de las operaciones mentales

Cuestiones sobre la naturaleza de los sistemas de procesamiento de información

¿Qué es la naturaleza de la generalidad?
Un único método general de búsqueda
El gran interruptor (experto)
Métodos débiles
Asamblea ejecutiva
¿Qué es un problema mal estructurado?

Conclusión

La primera parte de nuestro título. “Inteligencia Artificial”, designa un campo de actividad científica a través del cual haremos un breve recorrido. “El Concepto de Mente” designa, si bien de una manera poco clara, un grupo de problemas que debemos plantearnos. La discusión ha de versar sobre el modo y los medios en que los materiales diseminados a lo largo del referido campo científico contribuyen a clarificar, y tal vez hasta a solucionar, esa serie de problemas.

Tengo mucho más clara la naturaleza de la inteligencia artificial que los problemas que deben plantearse sobre el concepto de mente. Para determinar esto último, he decidido adoptar un punto de vista ortodoxo de la psicología científica: el de comprender a la mente por medio de la construcción de teorías (si se prefiere, modelos) sobre la conducta de la mente humana, comprobándolas por cualquier método empírico que parezca apropiado, y manteniendo después como “concepto actual de mente” la teoría que mejor consiga superar las pruebas. No se puede conseguir mucho más.

En esta discusión, describiré primero el campo de la inteligencia artificial, suponiendo que el conocimiento del lector es lo suficientemente impreciso como para que le beneficie tal intento. A continuación, formularé algunas cuestiones sobre el concepto de mente, e indagaré si la inteligencia artificial tiene algo que ofrecer para su solución.

LA NATURALEZA DE LA INTELIGENCIA ARTIFICIAL

Los campos científicos surgen cuando los intereses de los hombres de ciencia cristalizan en torno a diversos fenómenos. A las ciencias no se las define, sino que se las reconoce. Los actos de definición, cuando ocurren, son en realidad prescriptivos. Representan intentos para desviar el curso de una ciencia, para reclamar nuevos fenómenos, para delimitar la ciencia, o para demostrar que se la denomina erróneamente (un ejercicio usualmente inútil de lingüística aplicada).

Sin embargo, la inteligencia artificial es una parte todavía no cristalizada de la ciencia del computador, ciencia esta última que sólo recientemente ha alcanzado a cristalizar. Los fenómenos centrales son procesos de información que realizan tareas que exigen inteligencia. Dentro del dominio de tales tareas, aunque están delimitadas, diversos subdominios han recibido atención mientras que otros han sido virtualmente ignorados, y esto ha conducido a una pluralidad de concepciones sobre la naturaleza de la inteligencia artificial. Aunque estas concepciones no forman una totalidad integrada, su diversidad no es ilimitada, y cada uno puede establecer la suya propia como una forma de proporcionar sentido al estudio de la materia.

Me propongo presentar tres de estas concepciones o perspectivas de la inteligencia artificial. Cada una de ellas difiere un tanto de las demás, aunque naturalmente todas deben reflejar aspectos comunes de un mismo estado actual. Entre las tres, abrirán a nuestro escrutinio una gran parte del campo, y aunque algunas otras concepciones estarán visiblemente ausentes, creo que son las que menos tienen aquí que ofrecernos.

PRIMERA CONCEPCION:

LA EXPLORACION DE FUNCIONES INTELECTUALES

Supóngase que convenimos en que la cuestión principal de la inteligencia artificial es la siguiente: "¿Qué mecanismos pueden realizar determinadas funciones intelectuales?" En consecuencia, el desarrollo de la inteligencia artificial se convierte en una exploración donde cada función es abordada y conquistada —o quizá se resiste a todos los intentos de mecanización.

No hace falta disponer de un mapa preciso de la geografía de las funciones mentales, del mismo modo que no es necesario disponer de antemano de un mapa preciso de la geografía de un nuevo continente. Las dimensiones de cada valle y de cada montaña han de ser recorridas, exploradas y establecidas; la exploración podrá llevarnos después al interior de una nueva región.

Para que la analogía geográfica fuese válida, debería existir un espacio de funciones intelectuales que se pudiera definir con independencia de cualquier intento de explorarlo. No es esto, naturalmente, lo que ocurre. Pero existe, sin embargo, una rica noción informal de la diversidad de funciones intelectuales y de las relaciones que éstas guardan entre sí. Dichas funciones derivan de una gran variedad de fuentes —de nuestro lenguaje, de nuestras tradiciones filosóficas, de las ciencias de la psicología, la lingüística, la sociología, la biología, de la medicina, de la tecnología y de nuestra tradición literaria. Estas fuentes no son separables, puesto que se alimentan una de otra. Con todo, no es difícil enumerar una multiplicidad de funciones intelectuales y señalar las tareas que, según consenso general, parecen incluirlas.

De las tres concepciones que vamos a presentar, ésta es la más cercana a la realidad histórica, en términos de las motivaciones e inclinaciones de los hombres que han roturado el campo de la inteligencia artificial. ¿Qué panorama ofrece el campo, contem-

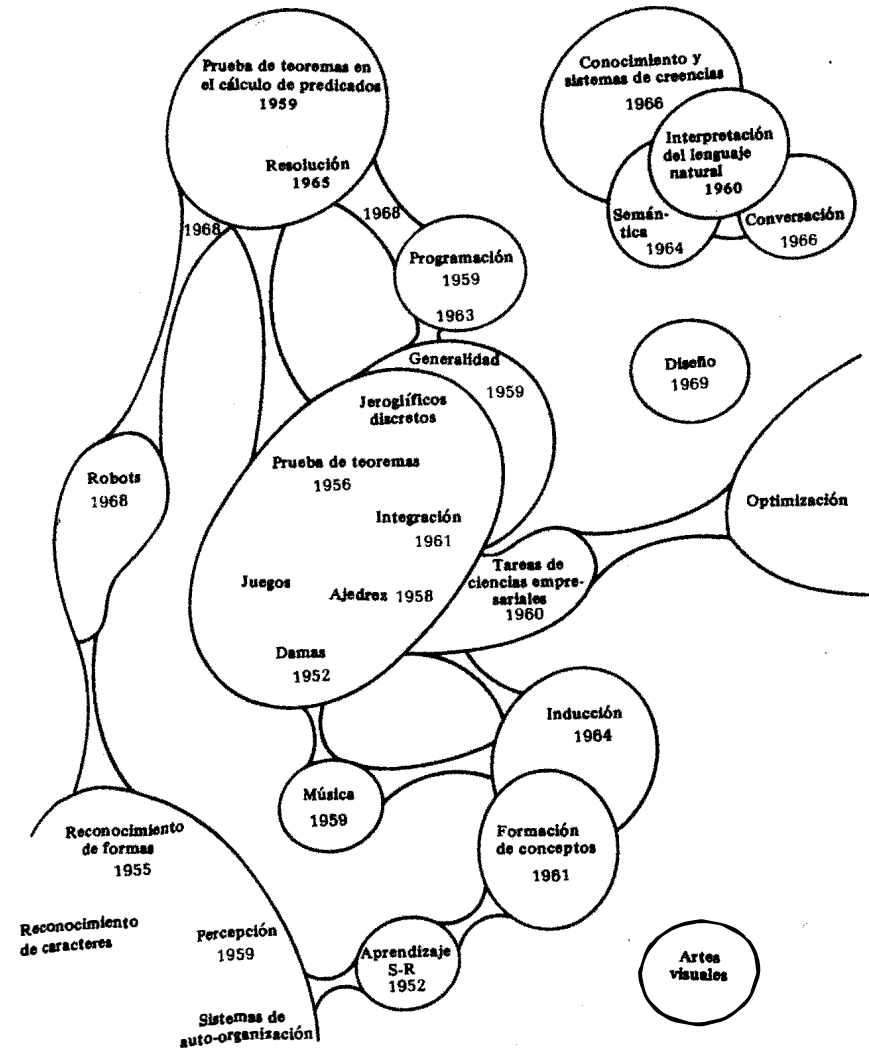


Figura 1.1 Geografía de la Inteligencia Artificial

plado de esta manera? Mi propia concepción actual de esta geografía se muestra en la Figura 1.1. Dibujar el mapa en dos dimensiones produce distorsiones, pero puesto que el territorio es metafórico, es preferible la fácil visualización a cualquier intento de representar con más exactitud la conexión entre las partes. Dentro de cada dominio, he introducido algunas fechas de esfuerzos característicos, incluyendo usualmente el primer trabajo que inició el área.

Esta geografía no lo es exactamente de las funciones intelectuales, sino más bien de las tareas que parecen incluir ciertas funciones intelectuales, aunque no siempre es fácil establecer cuáles son exactamente las funciones incluidas. (Con más frecuencia es posible determinar cuáles son las funciones que no están incluidas). Este giro o desplazamiento de la función a la tarea no se efectúa simplemente por conveniencia. Cuando un programa, por ejemplo, realiza el juego de ajedrez, es evidente lo que está haciendo, pero es menos claro si ese programa es de aprendizaje, de inducción, o de decisión. Para determinar si está realizando cualquiera de estas funciones, debemos ponernos de acuerdo con más precisión de la existente sobre el significado de estos términos. De esta manera, en última instancia, las funciones son problemáticas, y sólo la estructura de las tareas resulta ser el fundamento estable para describir lo que está haciendo un mecanismo.

Quisiera ser explícito sobre el significado que tiene para mí el término mecanismo (y, podría añadir, para el campo de las ciencias de la computación). Un mecanismo es cualquier proceso físico determinante. Un proceso abstracto constituye un mecanismo si, en principio, existen modos de realizarlo mediante un proceso físico. Así, cualquier programa de un computador digital constituye un mecanismo. De manera similar, una regla por medio de la cual podemos construir un artefacto físico que puede realizar su aplicación, es un mecanismo (o lo representa, si deseamos ser precisos). Esta idea puede ser formalizada mediante la noción de procedimiento efectivo, Máquina de Turing, Algoritmo de Markov, Sistema de Producción de Post. O podemos comenzar con los sistemas formales como noción primitiva (ideal) de meca-

nismo, y retroceder hacia los procesos físicos. Pero todo viene a ser lo mismo. La extensión del uso a mecanismos estocásticos, estadísticos o probabilísticos es directa, partiendo desde las nociones abstractas de probabilidad hasta los procesos físicos que obedecen a estos modelos formales.

Un recorrido de la geografía completa sería una empresa agotadora. [Una visión general de los diversos dominios puede encontrarse en dos libros recientes (NILSSON, 1971; SLAGLE, 1971)]. En lugar de ello, realizaremos un breve muestreo de cuatro áreas. Una incluye la programación, que da por supuesto un lenguaje de programación. Se especifican los datos de entrada y el resultado a obtener, y la tarea consiste en producir un programa. La salida, por tanto, constituye un programa realizado en un lenguaje de programación bien definido, y la entrada debe consistir en alguna especificación de lo que se desea. En el ejemplo PROW (WALDINGER y LEE, 1969), los programas están confeccionados en lenguaje LISP, y las especificaciones se realizan escribiendo una expresión de cálculo de predicados de primer orden de la forma:

$$\wedge x_1 \wedge x_2 \dots \wedge x_n \vee y_1 \dots \vee y_m P(x_1, \dots x_n, y_1 \dots y_m),$$

que debe interpretarse como “Deseo un programa que calcule las y s como salida para cualesquiera x s como entradas, donde las y s deben satisfacer el predicado P para las entradas dadas”. El programa de solución-de-problemas intenta resolver éste, planteándolo primeramente como un problema de prueba-de-teoremas —o lo que es lo mismo, probar el teorema anterior— que debe ser demostrado constructivamente, con ciertas restricciones sobre la demostración. A partir de esta prueba, si se la encuentra, resulta posible extraer un programa, satisfaciéndose así la tarea original.

Es evidente que la tarea a realizar aquí es una tarea intelectual. No es, desde luego, la misma tarea global que realizaría un programador humano. El programador comenzaría a partir de una representación más informal, como, por ejemplo, el lenguaje natu-

ral. Pero la transformación desde cualquier representación que pueda proporcionarse inicialmente a la que sirve de entrada para el ejemplo, no constituye aún la parte principal de las exigencias intelectuales de la tarea. Todavía queda por realizar la programación. El trabajo de GREEN (1969), que es muy similar al de Waldinger y Lee, está incorporado a un sistema con principio y final en lenguaje natural (COLES, 1969), realizando así efectivamente esta transformación.

El programa de Waldinger y Lee proporciona un mecanismo singular (el probador de teoremas) con el cual se realiza esta tarea. Este mecanismo no es necesario para la tarea de programar. Ciertamente, la novedad del programa de Waldinger y Lee radica en la reducción de una clase de tareas (programación) a otra (prueba de teoremas), donde la experiencia previa con la tarea inicial no sugiere una tal reformulación. La demostración de la suficiencia, más bien que la necesidad, de un mecanismo para una tarea es característica del trabajo empírico en inteligencia artificial.

La tarea no queda totalmente reducida a prueba de teoremas, ni siquiera con la eliminación de la traducción inicial. Aún falta la extracción del programa a partir de la prueba. A primera vista esto parece problemático, pero resulta posible escribir un algoritmo que realice siempre esa tarea, y sin excesivo esfuerzo. Así, la tarea intelectual reside, en su mayor parte, en la tarea de probar el teorema.

La segunda área consiste en la respuesta a una pregunta sobre una creencia mantenida por el programa. Dicho en términos más operacionales, un computador programado debe enfrentarse con preguntas tales como “¿Es un X un Y?”, y debe responder no sólo con “sí” o “no”, sino con las razones de su creencia. Existen varios programas que se ajustan al paradigma de tarea enunciar. Por mencionar uno, el programa de COLBY (1969) al tiempo de respuesta posee una red de información sobre el dominio que cae bajo su consideración, incluyendo hechos tales como “los niños agresivos son independientes”, y reglas tales como “si

un niño es agresivo, sus padres lo castigan”. Asociadas con estas estructuras simbólicas, hay algunas nociones, numéricamente definidas, de consistencia y credibilidad. La red se ha venido asimilando en el interior de la máquina durante anteriores interacciones con el mundo externo, que proporcionaron a la máquina tanto los datos como las preguntas que permiten al programa hacer inferencias a partir de su red y, de acuerdo con ellas, incrementar esta última. Por añadidura, el programa mismo formula una serie de interrogantes sobre su ambiente (un agente que interactúa con él) para obtener más información, modificando luego, en consecuencia, su red informativa.

De nuevo hay aquí, obviamente, una tarea intelectual, que puede ser razonablemente formulada diciendo que consiste en la construcción de un sistema de creencias y en la respuesta a preguntas sobre ese sistema de creencias.

La tercera área a la que voy a referirme es la de la solución general de problemas. Un programa de solución general de problemas consiste en resolver cualquier problema que pueda expresarse en un lenguaje formal especializado con algunas pretensiones de generalidad. Existen diversos programas que aceptan una tarea de este tipo, siendo uno de los más conocidos el GPS (ERNST y NEWELL, 1969), que acepta problemas que pueden plantearse en términos de objetos y de operadores que transforman esos objetos, e intenta solucionarlos recurriendo al análisis de medios-fines. Permítasenos, sin embargo, considerar un ejemplo más reciente, REF-ARF, construido por FIKES (1969). Su lenguaje (llamado REF) tiene la forma de un lenguaje de programación orientado-a-procedimiento-algebraico, aumentado por dos artificios que lo convierten en un lenguaje de enunciado-de-problema. En primer lugar, acepta funciones de la forma SELECT(1, N) que afirman únicamente que hay que seleccionar uno de los enteros entre 1 y N, ambos inclusive; en segundo lugar, acepta funciones de la forma CONDITION(...), que implica que la expresión escrita en el interior de los paréntesis debe ser tenida por verdadera. Un problema expresado mediante un programa escrito en este lenguaje

consistirá en hallar valores de las variables que hagan que termine el programa (con todas las condiciones satisfechas). Una persona que se proponga plantear problemas a REF-ARF tiene que traducir a estos términos su problema. Si lo consigue, entonces el sistema puede operar sobre el problema. El que pueda solucionarlo depende, naturalmente, de muchas cosas. Por ejemplo, el Cuadro 1.1 muestra un problema de criptoaritmética en inglés; el Cuadro 1.2 lo muestra escrito en REF; y el Cuadro 1.3 muestra la solución, tal como la imprime ARF (el solucionador de problemas), después de haber sido procesado el problema en un computador IBM 360/67. El ejemplo no sólo muestra el papel de las restricciones, sino también el uso del proceso lógico del programa para codificar la información del problema, puesto que el programa REF comprueba una solución sumando en la forma tradicional de derecha a izquierda, llevándose el exceso.

Hay una clara pretensión de generalidad en un programa tal como REF-ARF. Ello tiene lugar a dos niveles: ¿qué tipos de problemas pueden ser planteados en lenguaje REF, y qué clase de problemas expresables pueden ser resueltos en un tiempo razonable? Internamente, el programa ARF opera como un programa de búsqueda, de una manera muy similar a otros programas de juegos y de prueba de teoremas (lo que actualmente recibe el nombre de búsqueda o investigación heurística). Sin embargo, los nodulos de la búsqueda son problemas de satisfacción-de-restricciones, en lugar de simples posiciones de juego o teoremas. Así, ARF posee realmente dos conjuntos de procesos de solución-de-problemas, los que orientan la búsqueda heurística y los que intentan solucionar un problema particular en un nódulo particular. La rapidez con que ARF soluciona el problema en la Figura 3, se debe a las técnicas relativamente buenas de satisfacción-de-restricciones. Existen limitaciones de menor importancia a la generalidad de los lenguajes; por ejemplo, no está permitida la multiplicación. Pero también existen limitaciones de mayor importancia; por ejemplo, las variables son todas de enteros, de forma que las estructuras de datos tales como las listas con operaciones constructivas (inserción y eliminación) no están permitidas.

Cuadro 1.1 Problema de Criptoaritmética

C R O S S
+ R O A D S

D A N G E R

Cada letra representa un dígito, es decir, 0, 1, ..., 9. Cada letra es un dígito distinto, de forma que nunca se puede asignar el mismo dígito a dos letras diferentes. ¿Qué dígitos deberán ser asignados a las letras, para que cuando se reemplazaran las letras por sus correspondientes dígitos fuera correcta la suma?

Cuadro 1.2 Formulación REF-ARF del Problema de Criptoaritmética [De Fikes (200)]

```
BEGIN;
  SET VECTOR A1 TO X,C,R,O,S,S;
  SET VECTOR A2 TO X,R,O,A,D,S;
  SET VECTOR SUM TO D,A,N,G,E,R;
  SET VECTOR L TO C,R,O,S,A,D,N,G,E;
  FOR I → 9 DOTO L1;
L1: SET <I[<I>]> TO SELECT(0,9);
  CONDITION EXCI(<C>,<R>,<O>,<S>,<A>,<D>,<N>,<G>,<E>;
  CONDITION ~ (<C>=0) ^ ~ (<R>=C) ^ ~ (<D>=0);
  SET <CARRY> TO 0;
  FOR J → 5 DOTO L2;
  SET <I> TO 7+--<J>;
  IF <A1[<I>]>+<A2[<I>]>+<CARRY><10 THEN L3;
  CONDITION <A1[<I>]>+<A2[<I>]>+<CARRY>=10+<SUM[<I>]>;
  SET <CARRY> TO 1;
  GOTO L2;
L3: CONDITION <A1[<I>]>+<A2[<I>]>+<CARRY>=<SUM[<I>]>;
  SET <CARRY> TO 0;
L2::
  CONDITION <D>=<CARRY>;
END;
```

**Cuadro 1.3 Solución REF-ARF al Problema de Cripto-
aritmética [De Fikes (201)]**

SOLUTION FOUND AS FOLLOWS:	
CONTEXT 300205	
DATA STRUCTURE	
J	<J>: 6
CARRY	<CARRY>: 1
E	<E>: 4
G	<G>: 7
N	<N>: 8
D	<D>: 1
A	<A>: 5
S	<S>: 3
O	<O>: 2
R	<R>: 6
C	<C>: 9
I	<I>: 2
L	VECTOR: C,R,O,S,A,D,N,G,E
SUM	VECTOR: D,A,N,G,E,R
A2	VECTOR: X,R,O,A,D,S
A1	VECTOR: X,C,R,O,S,S

La cuarta área comprende la coordinación de sistemas de ojo-y-mano. Una tarea típica (véase FELDMAN *et al.*, 1971) consiste en construir un apilamiento con bloques que están inicialmente esparcidos. El rasgo distintivo de este problema consiste en que los bloques son reales —de tres dimensiones, sólidos, colocados en una mesa real, etc.— y que el computador posee un receptor visual (una cámara de TV) y un efector físico (una pinza). De esta forma, su problema es un problema de percepción y de consecuente coordinación de su poder perceptual con su poder efector. Tiene una representación interna de la tarea que es formal —esto es, simbólica y discreta. Pero esta representación es construida por el sistema mismo en el curso de su intento de resolver el problema, de forma que los errores surgen de las inadecuaciones que se insinúan dentro de tal representación interna.

La descripción de estas cuatro áreas proporciona algún contenido al género de territorio que ha sido y está siendo explorado bajo la rúbrica de inteligencia artificial. El desarrollo del campo durante los últimos veinte años se ajusta bastante bien a la concepción de la inteligencia artificial como exploración —en un desarrollo secuencial de una región tras otra. Sin embargo, también ha tenido lugar un desarrollo paralelo, principalmente en reconocimiento de formas, que ha recibido de continuo tanta atención en estos años como todas las demás regiones juntas, y en prueba de teoremas del cálculo de predicados. Distintas comunidades de investigadores han trabajado sobre cada región. El reconocimiento de formas ha sido siempre objeto separado de investigación, con el uso de matemáticas continuas, su inspiración en fisiología y biología, y su parcial independencia del computador digital. De la misma manera, la prueba de teoremas en el cálculo de predicados ha sido abordada independientemente por lógicos y matemáticos interesados en el desarrollo de la matemática mecánica, por usar el término de Hao Wang (WANG, 1960).

Al exponer la presente concepción, me he guardado, deliberadamente, de tratar de extraer un sentido de la plétora de tareas. En la Figura 1.1 las represento en yuxtaposición sólo cuan-

do parece obvio su parentesco. Por ejemplo, la inducción y la formación de conceptos se superponen, al igual que los diversos intentos de tomar contacto con el lenguaje natural, que aparecen en la esquina superior derecha.

No es difícil pensar en tareas que exigen actividad intelectual y que no están representadas. Elijamos cualquier aspecto de la vida humana —como, por ejemplo, el teatro— y detengámonos momentáneamente a considerarlo: escribir un drama, crear un diseño y un escenario, dirigir una obra teatral, ser gerente de una compañía (obtener todas las cosas que se necesitan y asegurarse de que estén disponibles y en condiciones adecuadas), improvisar un nuevo escenario, evaluar el efecto de un chiste en el auditorio, etcétera. Muchas de estas tareas poseen aspectos diferentes de los puramente intelectuales, pero cada una podría constituir el núcleo de una tarea purificada, a cuyo respecto podríamos preguntar después por los mecanismos intelectuales que la realizan. Y para cada una, existen dimensiones cualitativas; una cosa es realizar la tarea, otra realizarla pasablemente, y aún una tercera realizarla soberbiamente.

Existen, como es natural, formas de concebir la inteligencia artificial que no están tan exentas de teoría —que no se limitan a ver este campo como una sucesión de arduas tareas. Veremos ahora una de esas otras concepciones.

SEGUNDA CONCEPCION:

LA CIENCIA DE LOS METODOS DEBILES

Por lo que llevamos dicho hasta el presente, un programa es un bloque compacto de información codificada, creado para realizar una tarea de interés. (La tarea, naturalmente, puede tener sus propios aspectos de generalidad, como en REF-ARF). Es decir: una tarea, un programa —los programas que han sido mencionados no son capaces de realizar tareas que no sean las suyas propias. Constituyen entidades distintas.

Pero si existiese un pequeño número de ideas tras la multitud de programas individuales, entonces una descripción en términos de esas ideas podría poner de manifiesto una dosis bastante mayor de coherencia y de norma. Ciertamente, esto parece ser lo que ocurre. Parece, aunque no se dispone todavía de ninguna demostración concluyente, que estas ideas pueden ser recogidas en forma de *métodos*. La inteligencia artificial puede ser concebida, por tanto, como un campo dedicado al descubrimiento y recolección de un conjunto de métodos. De esta manera comporta un parentesco con campos tales como el análisis numérico, la programación matemática, y probablemente hasta la ingeniería química. La naturaleza de la inteligencia artificial, dentro de esta concepción, se deduce de la naturaleza general de sus métodos. Pero antes de intentar una caracterización general, observaremos los métodos con algún detalle.

Por método, naturalmente, entendemos una receta o procedimiento para realizar una clase de tareas. Aunque el término exacto que se utilice no es importante, lo es el distinguir tres componentes de un método:

El enunciado del problema: un enunciado en algún lenguaje que describe la situación a la que se aplica el método, y el problema al que éste puede proporcionar solución.

El procedimiento: un programa (es decir, una secuencia condicional de acciones), basado en algún repertorio de operaciones primitivas.

La justificación: las razones por las que se piensa que si se satisfacen las condiciones del enunciado del problema, entonces el procedimiento debe alcanzar la solución.

Los Cuadros 1.4 y 1.5 muestran dos ejemplos simples de métodos. Uno es la fórmula típica para la solución de una ecuación de segundo grado. Suponemos que el usuario del método puede transformar la expresión dada en un procedimiento paso a paso. Si el usuario no sabe extraer raíces cuadradas, entonces este método no es válido para él, puesto que supone tal operación en el repertorio básico. De igual manera, si una tarea no puede ser caracterizada en el enunciado del problema, entonces, y de nuevo, este método, simplemente, no es aplicable. En el caso de la ecuación de segundo grado, se puede obtener la justificación completa substituyendo dentro de la ecuación las raíces por sus fórmulas. Esto no sería tan trivial si el procedimiento se hubiese dado, por ejemplo, en código de máquina, aunque, sin embargo, no existiría ninguna dificultad fundamental.

El segundo método está tomado de un libro de reparación de muebles. Hay una considerable ausencia de rigor formal tanto en el enunciado del problema como en el procedimiento, pero la operación básica —sacar la puerta al exterior— es presumiblemente conocida por el lector. Se pone un considerable cuidado en la

Cuadro 1.4 Método para solucionar ecuaciones de segundo grado

Información dada:	números reales a, b, c , variable real x
Formulación del problema:	Hallar un x tal que $ax^2 + bx + c = 0$
Procedimiento:	Calcular $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$
Justificación:	Substituir x por su fórmula en $ax^2 + bx + c$ y mostrar por manipulación algebraica que el resultado es 0.

Cuadro 1.5 Método para enderezar una puerta de madera que está curvada

Puerta curvada. Este es el punto de partida para introducir una discusión general sobre las curvaturas —puesto que las puertas de madera son las víctimas más comunes. Para empezar, la curvatura se produce por la humedad que penetra en un lado de la lámina, sin poder hacerlo en el otro por impedirlo algún tipo de cubierta aislante. Si esto le sugiere que la forma de evitar la curvatura consiste en recubrir ambos lados de la puerta así como los remates, estamos completamente de acuerdo. De hecho, en un trabajo bien acabado esto es lo que se hace siempre.

Pero volviendo a la situación de la puerta que ya está curvada, el tratamiento consiste en humedecer el lado seco. Este lado es siempre el que está cóncavo (hundido). Afortunadamente, cualquier madera que ya se hubiese curvado se enderezará al realizar esta operación —con independencia de la antigüedad de la pieza o de la duración que tenga la curvatura.

Vamos a ilustrar el método más fácil para que la humedad penetre en el lado hundido de una puerta curvada. Coloque ese lado hundido o cóncavo sobre el césped, y deje que el sol caliente el otro lado. Esto puede tardar desde cuatro horas hasta cuatro largos días de sol. Cuando la puerta parezca haberse enderezado, métala en casa y póngala en una habitación de temperatura equilibrada durante unos cuantos días. Algunas veces vuelve a reaparecer, días después, un poco de curvatura. Si esto sucede, póngala de nuevo sobre el césped para realizar otra terapia más. De hecho, esta vez puede dejarla hasta que el lado que mira al sol llegue a adquirir una cierta concavidad. Esta no es una tarea peligrosa ni inquietante, porque la madera es increíblemente flexible. Se puede curvar la puerta primero de una parte, y luego de la otra, y volver a hacerlo sin estropear la madera.



descripción de comprobaciones sensoriales. Sólo se proporciona una breve justificación del procedimiento. Es en parte científica y en parte empírica. Se podría decir bastante más, naturalmente, pero es evidente que no se podría proporcionar ninguna justificación formal.

Demostrar que el trabajo en inteligencia artificial tiene el carácter de descubrimiento y desarrollo de un conjunto de métodos, requiere lo siguiente:

1. Mostrar un conjunto de métodos —es decir, los métodos de inteligencia artificial desarrollados hasta la fecha.
2. Demostrar que son métodos: proporcionar enunciados de problemas, procedimientos y justificaciones.
3. Demostrar cómo se pueden escribir programas efectivos de inteligencia artificial en términos de estos métodos.
4. Establecer que, en cierto sentido, la parte del programa total representada por el método proporciona una fracción sustancial (cuanto más, mejor) de la potencia total del programa.
5. Demostrar que los mismos métodos son utilizados en más de un programa, reduciendo así la diversidad de programas independientes en el campo.
6. Demostrar que los métodos abarcan una parte sustancial del territorio total de la inteligencia artificial.

Tal empresa no puede ser llevada a cabo por completo en un escrito como el presente. En otro lugar, y en un contexto diferente, defendí la concepción de la inteligencia artificial como la ciencia de los métodos débiles, suministrando algunos detalles más de los que puedo aducir aquí (NEWELL, 1969). Pero aun así, esta concepción no puede ser, de hecho, desarrollada a un nivel técnico adecuado. En cualquier caso, mi propósito actual es describirla, no demostrarla. Por tanto, me limitaré a resumir e ilustrar la mayoría de los puntos que acabo de mencionar.

La Tabla 1.1 proporciona una lista inicial de los métodos de inteligencia artificial. Las descripciones son breves, pero espero que permitan captar lo esencial. Se pueden encontrar más detalles

en NEWELL (1969) y existe una tabla de características de una serie de programas de búsqueda heurística en NEWELL (1965). La ubicuidad de alguno de estos métodos, tales como el de Ascensión Máxima (*Hill Climbing*) o el de Búsqueda Heurística (*Heuristic Search*), está ampliamente reconocida. Otros, como el de Construcción de Hipótesis y Comparación (*Hypothesize and Match*) están menos divulgados.

Tabla 1.1 Descripción de los Métodos de Inteligencia Artificial

Enunciado del problema	Procedimiento
Generar y Comprobar	
Dado: un conjunto X con generador; una propiedad P con prueba R.	Generar los elementos de X con G. Comprobar cada P con R.
Encontrar x en X tal que P(x)	Si no hay más elementos de X(G—), fracaso.
Búsqueda Heurística (versión simple)	
Dado: un conjunto X con operadores $Q(q:X \rightarrow X)$; una propiedad P con prueba T.	Mantener un elemento actual, x, (inicialmente x_0). Seleccionar un operador de Q y aplicarlo a x (si no se puede, seleccionar otro).
Encontrar una secuencia de operadores que transformen un elemento inicial x_0 en un elemento x tal que P(x).	Comprobar el nuevo elemento para P con T, si + el nuevo elemento es la solución. Colocar el nuevo elemento en "lista de prueba". Seleccionar el nuevo elemento actual de la lista de prueba.
Ascensión Máxima (versión simple)	
Dado: un conjunto X con operadores Q; una ordenación en X con proceso de comparación C; X es unimodal.	Mantener un elemento actual, x (seleccionar una x inicial). Seleccionar un operador de Q y aplicarlo a x (si no se puede, seleccionar otro).
Hallar el elemento más alto en X.	Comparar el nuevo elemento con el elemento actual mediante C. Si es más elevado, tomar el nuevo elemento como el actual. Si no es más elevado, descartarlo y seleccionar otro operador. Si no quedan operadores por probar el elemento actual es la solución.

Tabla 1.1 (Continuación)

Enunciado del problema	Procedimiento
Comparación (versión simple)	
Dado: un conjunto X que consta de partes; un conjunto de formas sobre X con operadores $Q(q:F \rightarrow FU X)$; una comparación C entre las partes que produce diferencias (en un conjunto D), tal que si una diferencia aparece entre dos partes, entonces existe a lo sumo un operador que puede "eliminarla".	Generar partes correspondientes del elemento y la forma. Comparar las partes utilizando C. Si no hay diferencia, continúa. Si hay diferencia, entonces selecciona un operador que elimine la diferencia. Si no existe, entonces fracaso. Aplicar operador, si hubo éxito continúa, en caso contrario fracaso.
Hipótesis y Comparación	
Dado: un conjunto de hipótesis H, cuyos elementos son funciones de un dominio D a otro dominio o rango $R(h:D \rightarrow R)$; un conjunto de formas F sobre H; un conjunto de ejemplares E de la función $(d \rightarrow r)$. Descubrir una hipótesis que satisfaga a todos los ejemplares en E.	Generar las formas yendo de lo simple a lo complejo. Para una forma dada, generar los ejemplares. Para cada ejemplar aplicar la forma a su dominio produciendo un elemento esperado del rango r' (que puede ser una forma). Comparar r' con el elemento del rango del ejemplar, con operadores que instanciarán r' (y mediante él, la forma). Si la comparación fracasa, entonces la forma fracasa; obtener la siguiente. Si la forma tiene éxito para todos los ejemplares, entonces (así instanciada) constituye la solución.

Para mostrar las características principales de estos métodos, examinaremos el más simple de todos ellos: Generar y Comprobar. El Cuadro 1.6 proporciona un desarrollo explicativo de ese método. En primer lugar, hay un enunciado del problema que contiene cuatro variables. Obsérvese que dos de ellas, G y T, se refieren a procesos. Es decir, para utilizar este método hay que disponer de ciertos subprocesos, y el método indica cómo organizarlos para alcanzar una cierta meta, que es, concretamente, la que viene determinada por el enunciado del problema. Lo que hay que hallar es definido en términos de X y de P, que no son procesos. Por tanto, es necesario que exista alguna conexión entre los procesos G y T, que pueden ser ejecutados dentro de alguna estructura ejecutiva, y X y P. Esta conexión viene proporcionada en una primera instancia por las definiciones de G y T, que relacionan estos procesos con lo que hay que hacer, y después por enunciados específicos que aseveran que G ha de generar el conjunto S, y que T ha de comprobar el predicado P.

El procedimiento es bastante simple: la salida del generador debe convertirse en la entrada para la comprobación; si la comprobación tiene éxito, se llega al final denominado "solución"; en caso contrario, el generador debe producir otro elemento; si el generador se detiene alguna vez, entonces se llega al final denominado "fracaso".

La justificación es similarmente simple y aquí sólo está esquematizada. Para exponerla rigurosamente, se necesitaría formalizar el concepto de proceso, cosa que se ha realizado y se puede realizar sin dificultad. Hay que probar dos proposiciones: que si terminamos en el punto marcado como *solución*, entonces tenemos realmente una solución; y que si terminamos en el punto marcado como *fracaso*, entonces no existe solución al problema. Estas proposiciones justifican en algún grado el uso del método. No proporcionan toda la justificación que uno podría desear —por ejemplo, el esfuerzo esperado para obtener una solución, o una medida en el tiempo de las probabilidades de obtenerla. Tales característi-

Cuadro 1.6 Método de Generar y Comprobar

Información dada:

X es un conjunto.

G es un generador para el conjunto X.

P es un predicado definido sobre X.

T es un comprobante de P.

T es un proceso de comprobación si:

Output (+) implica que el output es input y P(input).

Output (−) implica que no P(input).

G es un generador si:

Output (+) implica que el output está en X.

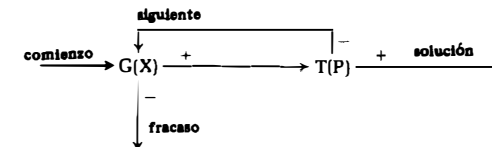
Output (−) implica que todos los elementos han sido generados; i.e., ocurrieron en alguna ocasión como output(+).

Input (comienzo) inicia la acción para producir el siguiente output.

Enunciado del problema:

Hallar un x tal que x esté en X y P(x).

Procedimiento:



Justificación:

Hay que mostrar: (1) Terminar en solución implica que el enunciado del problema está satisfecho.
(2) Terminar en fracaso implica que no hay un x tal que x esté en X y P(x).

1. Terminar en solución con x implica $x = T\text{-output}(+)$, implica $P(x)$ y $x = T\text{-input}$, implica $x = G\text{-output}(+)$, implica que x está en X, implica (1).
2. Terminar en fracaso implica que nunca T(x) para $x = T\text{-input}$; terminar en fracaso implica que todo lo generado en X por G es G-input(+), implica que todo x en X es input a %, implica que para ningún x en X P(x), implica (2).

cas no están determinadas por la formulación dada, puesto que se necesitaría más información sobre la naturaleza de G y T. Por ejemplo, si sabemos que el conjunto es finito, entonces sabemos que eventualmente el procedimiento debe terminar y que el método es un procedimiento de decisión para el problema.* En realidad, lo que necesitamos es una finitud relativa a los recursos de computación disponibles, puesto que “finito, pero inalcanzable” es casi tan insatisfactorio como “no finito”.

Los métodos de la Tabla 1.1 pueden ser desarrollados en enunciados del problema y procedimientos al estilo del Cuadro 1.6, y ciertamente la totalidad resultante puede ser plasmada en un sistema formal. Todos esos métodos implican la especificación de algunos procesos además de las variables establecidas. Las justificaciones exigen algo más de atención, y ello por dos razones. En primer lugar, a diferencia del ejemplo de la ecuación de segundo grado, los métodos de inteligencia artificial están expresados en un lenguaje de procedimiento, y sabemos poco sobre cómo establecer su validez. Esto se pone de manifiesto, por ejemplo, al darnos cuenta de que lo que necesita un programa escrito para realizar una tarea no es tanto que se demuestre que es correcto como que se lo depure —es decir, que se lo someta a comprobación experimental con el subsiguiente descubrimiento y supresión de las dificultades. Puesto que un programa es un objeto simbólico formal, no existe ninguna razón por la que no se lo pudiese tratar matemáticamente. En esencia, así hemos hecho con el método de Generar y Comprobar, sólo que su simplicidad encubre las dificultades. Hay trabajos recientes de interés sobre cómo asociar con programas los enunciados descriptivos adecuados (usualmente proposiciones de un cálculo lógico) de lo que dichos programas deberían hacer, de modo que se puede demostrar si lo hacen. Se sabe, por ejemplo, que podemos formalizar la programación de tal modo que a cada programa corresponda un predicado de lógica de

* Esto supone el aislamiento del método respecto de las influencias externas, como por ejemplo el estar terminado. En un contexto más amplio, tales posibilidades complican los detalles de la prueba.

primer orden tal que el programa termina, y termina correctamente, si y sólo si este predicado es verdadero (véase FLOYD, 1967 y MANNA y VUILLEMIN, 1972). El ejemplo de Waldinger, discutido anteriormente, refleja esta corriente de investigación (véase también MANNA, 1969).

La segunda razón concierne a lo que hay que justificar. La mayoría de los programas de inteligencia artificial no son algoritmos completos para las tareas que se proponen solucionar (esta es una de las razones por las que prefiero llamarlos por otro nombre: “métodos”). E incluso cuando pueden ser convertidos en algoritmos mediante expedientes sencillos, como ha podido observarse en Generar y Comprobar, muchos de ellos no son algoritmos prácticos. Las demostraciones deseadas dicen relación a cosas tales como el esfuerzo esperado, el riesgo de quedar atrapados sin ulterior probabilidad de hallar una solución, etc. Más aún, precisamos de tales evaluaciones como una función del ambiente que circunda a la tarea, lo cual implica de suyo la introducción de conceptos apropiados para describir ese ambiente.

En realidad, la teoría sustancial de este tipo que ha podido formularse para los métodos de inteligencia artificial, es escasa. Los dos más notorios contra-ejemplos a este aserto son intentos de demostrar la completud de numerosas variaciones de técnica de búsqueda en la prueba de teoremas del cálculo de predicados (véase ROBINSON, 1967), y la teoría de los perceptrones, tal y como ha sido recientemente desarrollada por MINSKY y PAPER (1969). En el primer caso, se desea averiguar si al limitar la búsqueda a un camino particular se ha eliminado toda posibilidad (para algunos problemas) de hallar la prueba. En general, nada dice esto sobre los relativos desplazamientos de los esfuerzos esperados para hallar pruebas. Sin embargo, la motivación para introducir estos nuevos métodos consiste usualmente en eliminar alguna manifiesta ineficiencia observada en los métodos de búsqueda existentes. Así, las principales justificaciones del método descansan en la plausibilidad de eliminar la ineficacia y en la demostración empírica sobre un reducido conjunto de tareas.

La investigación sobre perceptrones ha llevado a dos tipos

de resultados. Uno está relacionado con el hallazgo de un perceptrón que ha de realizar una tarea especificada, por medio de diversas formas elementales de modificación adaptativa, supuesto que un posible perceptrón exista (el Teorema de Convergencia del perceptrón). El otro resultado especifica diversas subclases de perceptrones y proporciona pruebas de que existen miembros de estas subclases que pueden realizar diversas tareas de reconocimiento (o que ninguno puede hacerlo). Un resultado típico consiste en que ningún perceptrón limitado en orden o limitado en diámetro (dos formas de limitar un perceptrón a medidas "locales" sobre su retina) puede discriminar la clase de figuras conectadas de las no conectadas.

La discusión anterior hace plausible la noción de un conjunto de métodos de inteligencia artificial sólo en relación a los dos primeros puntos y al último de la lista presentada en la página 20: mostrar un conjunto de métodos con sus enunciados del problema, procedimientos y justificación, y demostrar que abarcan al menos una porción del campo total de la inteligencia artificial. De qué manera la abarcan ha sido indicado sólo por las referencias contenidas en la Tabla 1.1. Existen diversas áreas que han sido omitidas, principalmente el área de los robots y la de las grandes redes semánticas y sistemas de creencia.

En cuanto a los demás puntos de la lista, que tratan de la descripción de un programa real en términos de métodos y el uso del mismo método en distintos programas, consideraremos una sola ilustración.

Hipótesis y Comparación

Las tareas inductivas han sido siempre una parte prominente del paisaje de la inteligencia artificial. Las razones al respecto parecen ser dos. Por un lado, hemos heredado una distinción clásica entre deducción e inducción, de suerte que la investigación sobre la acción inteligente se orientaría claramente hacia la inducción. En segundo lugar, la psicología americana ha identifica-

do en gran medida el problema central de la conducta conceptual con la adquisición y formación de conceptos —que en la práctica ha venido a significar la inducción de conceptos a partir de un conjunto de ejemplares presentados. Esta tendencia, que acusa el fuerte influjo de BRUNER, GOODNOW y AUSTIN en su *Study of Thinking* (1956), se deriva fundamentalmente del énfasis sobre el aprendizaje que ha caracterizado a la psicología americana desde el surgimiento del conductismo.

Aquí entran en juego los métodos utilizados en los distintos programas para realizar estas tareas. La tesis a establecer es que hay sólo unos pocos métodos (de acuerdo con el punto 5 de la página 20) y que éstos permiten explicar una parte sustancial de la potencia total en solución-de-problemas que tienen los programas.

Consideremos las tres tareas descritas en el Cuadro 1.7: una tarea típica de formación-de-conceptos, una tarea de extrapolación-de-secuencia y una tarea de analogía. Todas estas tareas tienen algo en común: se presentan ejemplares y la meta consiste en formar una inducción, que puede ponerse de manifiesto en la creación, selección o uso de otros ejemplares. Los ejemplares mismos constituyen la única información a partir de la cual se elaboran las construcciones (o selecciones) finales, lo que justifica el rótulo de inductivas para estas tareas. Otro rasgo común de dichas tareas es que los ambientes que las circundan son codificables en estructuras simbólicas discretas (admitiendo la geometría analítica para la tarea analógica).

Para cada tarea hemos tomado en consideración un programa: el de JOHNSON (1964) para la formación de conceptos, el de SIMON y KOTOVSKY (1963) para la extrapolación de secuencias, y el de EVANS (1964) para las analogías geométricas. Parece que el método singular que hemos llamado Hipótesis y Comparación contiene la potencia esencial en solución-de-problemas de todos ellos. Evidentemente, los aludidos programas no son los únicos que realizan esas tareas. Muchos otros utilizan también esencialmente el mismo método, aunque algunos proceden de forma bastante distinta (véase HUNT *et al.*, 1966).

Cuadro 1.7 Tres tareas de Inducción

Tarea de formación de conceptos

0 0 0 0 0 rojo
0 0 0 0 0 rojo
0 0 0 0 0 verde
0 0 0 0 0 rojo
0 0 0 0 0 —
0 0 0 0 0 —

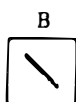
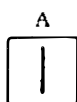
1. Para cada línea diga:
 - a. Si piensa que es un ejemplo de concepto-rojo o de concepto-verde.
 - b. Proporcione la hipótesis que crea válida para el concepto rojo (alternativamente, el concepto-verde).
2. Después se le dirá de qué concepto es un ejemplo la línea.
3. Después, repita para la siguiente línea.

Tarea de extrapolación de secuencia

M A B M B C M C D M — —

Complete los dos blancos para terminar la secuencia.

Tarea de analogías geométricas



La figura A es a la figura B como la figura C es ¿a cuál de las figuras?: D1, D2, D3, D4, D5.



El método de Hipótesis y Comparación se describe brevemente en la Tabla 1.1 (págs. 22-23) y con más detalle en el Cuadro 1.8. Es una subespecie del método Generar y Comprobar, con una característica especial que le proporciona su potencia. En el caso general, generamos una hipótesis después de otra, y comprobamos si funciona cada una de ellas —esto es, se aplica el método de Generar y Comprobar a un conjunto de hipótesis. Si el espacio de hipótesis es muy amplio —como ocurrirá en casi todos los casos de interés—, entonces o bien hay que añadir algo más o no se llegará a solucionar ningún problema. La característica especial del método de Hipótesis y Comparación es que genera no hipótesis particulares, sino representaciones de clases de hipótesis, de tal forma que los ejemplares pueden determinar directamente (esto es, sin una posterior generación) si una clase contiene una hipótesis que case con todos los ejemplares.

Para poner esto más en claro, observemos detalladamente el Cuadro 1.8, utilizando la tarea de extrapolación-de-secuencia para suministrar particulares. El método que aparece en el Cuadro 1.8 no muestra, ni mucho menos, la totalidad del programa, sino que es tan sólo la parte central del mismo, y tiene que adaptarse a cada tarea particular. El problema solucionado por el método consiste en hallar una aplicación (h) de un dominio (D) a otro dominio (R), dado que tenga una secuencia de ejemplares —esto es, pares particulares de elementos (d, r) tomados de D y R , de los que se sepa que satisfacen $h(d) = r$.

En las tareas típicas de formación-de-conceptos, la identificación natural consiste en que D está por los ejemplares presentados y R por el nombre del concepto [por ejemplo, (cuatro patas, con piel, y que dice guau-guau) → “perro”], y es así como se utiliza el método en los programas de Johnson. Para la tarea de secuencia, la identificación adecuada consiste en tomar D como el conjunto de subsecuencias iniciales y R como el conjunto de elementos. Es decir, la secuencia presentada MABMBCMDM proporciona los siguientes ejemplares: → M, M → A, MA → B, MAB → M, ... Como esta no es la única forma en que podría analizarse la tarea presentada, una parte de la labor de resolver el

Cuadro 1.8 Método de Hipótesis y Comparación

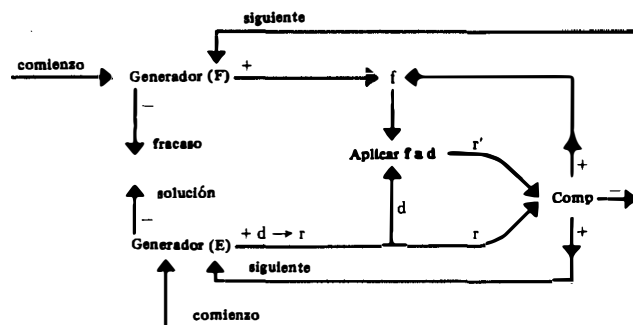
Información dada:

Un dominio $D = \langle d \rangle$
 Un rango $R = \langle r \rangle$
 Un espacio de hipótesis $H = \langle h \mid h: D \rightarrow R \rangle$
 Un espacio de formas $F = \langle f \mid f \text{ es una forma en } H \rangle$
 Un conjunto de ejemplares $E = \langle e \mid e = d \rightarrow r \rangle$

Enunciado del problema:

Hallar una hipótesis que se ajuste a todo E
 Hallar una h tal que para todo e en E $h(d_e) = r_e$
 (enunciado alternativo)

Procedimiento:



problema consiste en seleccionar o construir el conjunto de ejemplares, dadas la presentación y las instrucciones.

La cuestión crítica en el uso del método estriba en seleccionar el espacio de las hipótesis, es decir, la clase de aplicaciones. Al margen de los requisitos técnicos impuestos a la representación de este espacio por el método, que veremos más adelante, el éxito depende claramente de que se seleccione una clase de aplicaciones

que no sea tan grande como para no poder manipularla, pero que sea lo suficientemente grande como para que tenga una buena probabilidad de contener una respuesta al problema.* El salto inductivo crítico tiene lugar en este punto. Como si se tratase del truco de un prestidigitador, una vez se alcanza a dar ese salto desde el principio, el juego está realmente ganado. El resto se reduce a mera exhibición —lo que se traduce diciendo que el resto consiste en procesos “no inductivos” del tipo utilizado por los probadores de teoremas, jugadores de ajedrez, etc.

No es exactamente la clase de aplicaciones misma, sino sus representaciones lo que constituye el problema. Porque existen muchas formas distintas de representar la misma totalidad de aplicaciones. El programa de Johnson, por ejemplo, admite todas las funciones booleanas sobre D (que es un vector de valores booleanos). Estas funciones podían obtenerse mediante todas las expresiones lógicas formadas, por ejemplo, con y [conjunción], o [disyunción] y no [negación]; en lugar de ello, el programa admite conceptos tales como *simétrico* y *mayoría*. Esto determina en gran medida cuáles hayan de ser los conceptos primitivamente generados en la solución de problemas, puesto que el método de Hipótesis y Comparación considera invariablemente primero las aplicaciones más simples, tal y como vienen definidas por la representación.

La clase de aplicaciones para la extrapolación de secuencia se define por medio de una gramática, como la que se da en BNF (Forma Normal de Backus) en el Cuadro 1.9. Así, una expresión típica tal como

$$h: x \leftarrow A/\text{alf}[M \times n \times x]$$

puede leerse: existe una variable de indicador en el alfabeto, alf (esto es, $A, B, \dots, Z$), que debe ser inicializada en A ; después puede generarse la secuencia mediante un procedimiento cíclico,

* Muchos problemas engañosos de inducción se basan en un espacio oscuro. El ejemplo clásico en secuencias de letras es UDTCCSSO..., que se deriva de las letras iniciales de los números dígitos, de forma que el siguiente elemento sería N [nueve].

donde, en primer lugar, se imprime M , después se imprime x (en nuestro caso A), luego x pasa a ser la siguiente letra del alfabeto, a continuación se imprime de nuevo x (que es ahora B), y después se repite el ciclo (por tanto con M, B, C , después con M, C, D , y así sucesivamente). De esta forma, la información que se extrae sobre el dominio (hasta donde enseña la secuencia) consiste en una serie de indicadores que señalan a una serie de alfabetos. Se supone que la secuencia es periódica, con ciclos de n elementos. Cada elemento consiste en una secuencia de acciones, seguidas por la impresión de algún indicador. Las acciones especificadas son exactamente dos: mover un indicador al siguiente (n), y colocar un indicador en el lugar del otro ($x \leftarrow y$). Se pueden añadir otras acciones.

Para cada expresión del lenguaje existe una secuencia determinada. Por tanto, uno puede pedir de una expresión que esté de acuerdo con el conjunto dado de ejemplares. Es evidente que existe un conjunto ilimitado de posibilidades, tanto de expresiones como de las aplicaciones que éstas designen. La característica fundamental del método, tal como se indica en el Cuadro 1.8, consiste en generar no aplicaciones particulares, sino representaciones de *clases* de aplicaciones. En este caso, la clase de todas las aplicaciones con un período determinado viene producida por la introducción de variables sintácticas (simbolizadas por letras griegas) para las distintas partes de la expresión. Así, la forma $[a \beta]$ representa todas las aplicaciones de período dos.

La característica importante de esta representación consiste en que pueden utilizarse ejemplares para determinar directamente una aplicación que funcione (si es que existe). Esto sucede en dos etapas. La primera es que puede utilizarse la forma para producir algunos ejemplares. Estos ejemplares contendrán variables sintácticas —es decir, también ellos son formas. En la segunda, estas formas-de-ejemplares pueden ser comparadas con los ejemplares que se proporcionan como datos. El resultado de una comparación es la detección de una diferencia entre la forma-de-ejemplar y el ejemplar (si no hay diferencia, entonces la aplicación produce el ejemplar y todo queda en orden). Si estas diferen-

cias no pueden ser eliminadas, entonces no existe ninguna hipótesis con el período dado que produzca el conjunto dado de ejemplares. Si se las puede eliminar, la eliminación de una diferencia puede determinar alguno de los aspectos variables de la forma de ejemplar, y a través de este proceso, alguno de los aspectos variables de la aplicación. Así, paulatinamente, todos los grados de libertad en la aplicación se van tornando fijos hasta que emerge una aplicación específica.

La eliminación de diferencias puede representarse por una tabla en la que para cada diferencia entre la forma de ejemplar y el ejemplar se suministra un operador que modifique o especifique posteriormente la forma (y , a través de ella, la aplicación). Esta tabla de conexiones entre diferencias y operadores se muestra también en el Cuadro 1.9. Posee una importante propiedad en la que se basa la justificación para denominar al procedimiento *comparación*: si se ha hallado una diferencia de tipo d , entonces, si existe cualquier instanciación que obtenga éxito de la forma ejemplar, podremos hallarla aplicando el operador, $q(d)$. Por tanto, cuando nos enfrentamos con una diferencia d , no hay que hacer otra cosa sino aplicar $q(d)$. La posibilidad de que otros modos de manipular la forma pudieran conducir a una solución, no necesita ser en absoluto considerada. En particular, el método no necesita tener la capacidad de remontarse a buscar otras posibles secuencias alternativas de operadores para especificar la forma de ejemplar.

Cuadro 1.9 Espacio de Hipótesis y Tabla de Diferencias para la tarea de Extrapolación de Secuencias

Notación: BNF más $\text{sec-x} := x \mid x \text{ xsec}$
función $::=$ inicialización-sec ciclo
 nota: cada indicador del ciclo tiene una inicialización
ciclo $::=$ [elemento-sec]
elemento $::=$ indicador $\mid$ elemento de acción
acción $::=$ indicador $n \mid$ indicador $\leftarrow$ indicador
 nx cambia el indicador x por el siguiente en su alfabeto.
 $x \leftarrow y$ cambia x por lo que indique y .
indicador $::=$ $x \mid y \mid z$
alfabeto $::=$ $\text{alf} \mid \text{balf} \mid \dots$
 cada alfabeto es una secuencia particular de caracteres distintos.
carácter $::=$ $\dots$
inicialización $::=$ indicador $\leftarrow$ carácter/alfabeto.

Reglas de interpretación:

1. Interpreta cada elemento del ciclo en secuencia y después vuelve a empezar.
2. Para cada elemento, realiza las acciones en el orden indicado; cualquier acción de no-asignación o un indicador no asignado se consideran nulos.
3. El símbolo que esté a la derecha de un elemento es un indicador; imprime su valor actual. Si está sin asignar, utiliza la inicialización.

	Diferencia	Acción
Variable sintáctica (llámese α)	ϕ	$\alpha \leftarrow$ nuevo indicador
	nx	$\alpha \leftarrow nx \ x$
Valor de Indicador (desde x)	ϕ	fracaso
	nx	si elemento = x entonces $x \leftarrow nx \ x$, en caso contrario fracaso
	no def	añadir inicialización; $x \leftarrow$ carácter

El Cuadro 1.10 muestra la actuación de este esquema, siguiendo el diagrama de flujo del Cuadro 1.8. Utilizamos un generador simple de formas que comience con el período 1, después siga con el período 2, y así sucesivamente, lo que viene a ser la ejecución del principio "primero las aplicaciones más simples". El rastreo de una forma de ejemplar dada se mueve de izquierda a derecha a través de la página.* Cada diferencia, cuando ocurre,

Cuadro 1.10 Conjuntos de Movimientos Oculares y Conducta Verbal

	M	A	B	M	B	C	M	C	D	M	-	-
$[\alpha]$	$\alpha \Rightarrow \phi$											
$[x]$	x no def											
$x \ M \ [x]$	$M \ M \Rightarrow \phi$											
$[\alpha \ \beta]$	$\alpha \Rightarrow \phi$											
$[x \ \beta]$	x no def											
$x \leftarrow M \ [x \ \beta]$	$M \ \beta \Rightarrow \phi$											
$x \leftarrow M \ [x \ y]$	$M \ y \Rightarrow$ no def											
$y \leftarrow A, \ x \leftarrow M \ [x \ y]$	$M \ A \ M \Rightarrow \phi$											
$[\alpha \ \beta \ \gamma]$	$\alpha \Rightarrow \phi$											
$[x \ \beta \ \gamma]$	$x \Rightarrow$ no def											
$x \leftarrow M \ [x \ \beta \ \gamma]$	$M \ \Rightarrow \phi$											
$x \leftarrow M \ [x \ Y \ \gamma]$	$M \ y \Rightarrow$ no def											
$y \leftarrow A, \ x \leftarrow M \ [x \ y \ \gamma]$	$M \ A \ y \Rightarrow ny$											
$y \leftarrow A, \ x \leftarrow M \ [x \ y \ ny \ y]$	$M \ A \ B \ M \ B \ C \ M \ C \ D \ M \ D \ E$											

* Una ejecución a través de la página re-codifica los ejemplares separados en la continua predicción del siguiente elemento de la secuencia. De hecho, el diagrama de flujo del Cuadro 1.8 debe modificarse para trabajar a partir de un generador de elementos-siguientes, que construye el siguiente ejemplar total reteniendo la secuencia previa y tomando al nuevo elemento como el elemento del rango. Puesto que la información to-

termina un recorrido a través de la secuencia; la acción a realizar (o el fracaso) se escribe a la derecha de la flecha, y la forma de ejemplar nuevamente modificada (si es que existe) comienza una nueva ejecución. Así, considerando la primera línea, que proviene de la aplicación de período 1, el primer elemento es la variable sintáctica, α , puesto que aún no se ha especificado nada. Esto produce una diferencia, con el primer ejemplar (M) dirigiendo al operador, que sustituye un elemento de acción por una nueva variable indicadora (en este caso x). La siguiente ejecución produce una diferencia entre una variable no definida, x , y A , que lleva a establecer el enunciado de iniciación, $x \leftarrow A$ (aunque dejando aún sin definir el alfabeto). Finalmente, la tercera ejecución conduce a una diferencia para la que no existe operador correctivo, y por tanto, la forma es descartada y se genera la forma siguiente, $[a \beta]$.

Puede verse en el Cuadro 1.10 que la solución correcta es obtenida cuando se genera finalmente una forma del período correcto (3). El programa garantiza la operación del propuesto por Simon-Kotowsky, con la sola diferencia de que este último utiliza una investigación inicial para estimar el período de la forma de ejemplar y producir así únicamente el rastreo del período 3. Dicho programa, naturalmente, no está ni mucho menos organizado de la misma forma que el esquema que aquí se presenta. Nuestra refundición de los programas de inteligencia artificial en un número limitado de métodos es una reconstrucción racional.

No nos detendremos aquí en el análisis de los otros dos programas, pero tienen dos características esenciales que deben mencionarse. En primer lugar, abordan el problema de la induc-

tal extraída del elemento del dominio es el estado de los indicadores, no existe necesidad alguna de crear ninguna representación directa del elemento siguiente de D —todo el trabajo ya se ha realizado. La exigencia de adaptar el método a estas particularidades de la tarea de extrapolación de secuencia no debiera ser ignorada al evaluar si un método determinado cubre un programa particular de solución de problemas, pero esta cuestión guarda un paralelo con la de la compilación ineficiente. Es decir, podríamos proceder a crear cada ejemplar *de novo* e introducirlo en el proceso de aplicar —o lo que es lo mismo, utilizar el método no adaptado del Cuadro 1.8. Encontraríamos entonces (para nuestro disgusto como programadores) que se estaba repitiendo una cierta cantidad de cálculos.

ción disponiendo de un gran espacio de hipótesis —esencialmente, todas aquellas que son expresables en un lenguaje especializado. En ambos programas puede hacerse esto en BNF, al igual que sucede en el programa de Simon-Kotowsky. En segundo lugar, dividen su exploración del espacio de hipótesis en dos partes: la generación de subclases de hipótesis por medio de formas, y la selección de hipótesis particulares dentro de las posibilidades de cada forma por medio de comparación. El aspecto de generación raramente lleva asociadas heurísticas de interés, y puede implicar la generación combinatoria de todas las posibilidades.

En el programa de Johnson, una hipótesis simple sería: “rojo: columna 1 = negro”. Cuando se lo genera como una forma, el tipo de concepto (o rojo o verde) es especificado al igual que la columna, pero se deja variable el valor: “rojo: columna 1 = α ”. Esto se compara luego con los ejemplares. El programa de Johnson tiene otros dos métodos. La generación de la forma (la clase de hipótesis) se realiza por un mecanismo probabilístico (tomando decisiones aleatorias entre las diversas opciones gramaticales). Este mecanismo está sujeto a aprendizaje estocástico, que es esencialmente el de Subida Máxima sobre la frecuencia pasada de éxito. El otro, que puede llamarse método ptolomeico, pone retroducción de nuevos ejemplares. Las aumenta con disyunciones o conjunciones que o bien amplían, o bien reducen la hipótesis en cuestión para hacer frente a los ejemplares discrepantes. No es necesario decir que este método conduce a hipótesis complejas y esotéricas.

En el programa de Evans, la forma de la tarea dicta una adaptación más sustancial del método básico que en los otros dos. Su lógica es la siguiente:

Hallar una aplicación T que satisfaga $A \rightarrow B$.

Hallar aplicaciones que satisfagan $C \rightarrow D_i$, donde $i = 1, 2, 3, 4, 5$.

Abstractar (T, T_i) para obtener aplicaciones T_i' consistentes con T .

Seleccionar la más concreta de las T_i' .

También son posibles otras lógicas; la anterior evita la compara-

ción de objetos complejos (las figuras) o aplicaciones. Una parte de la inteligencia del programa radica en esta lógica, pero otra gran parte radica en las seis inducciones realizadas para hallar las aplicaciones. El aspecto de las aplicaciones que permite la comparación (y por tanto, la aplicación del método de Hipótesis y Comparación) consiste en la determinación de si un objeto (por ejemplo, un triángulo) puede transformarse en otro objeto (por ejemplo, un cuadrado) por medio de un cambio de escala, rotación y reflexión. Una vez que estas posibilidades son conocidas para todos los objetos en dos figuras (por ejemplo, en A y en B), el resto de la aplicación se puede coordinar por medio del método de Generar y Comprobar.

Métodos Débiles

Hemos considerado con algún detalle, aunque no por completo, la idea de que la inteligencia artificial puede ser concebida como el descubrimiento de un conjunto de métodos particulares, de forma que los resultados que podría alcanzar a la larga serían similares, por ejemplo, a los del análisis numérico: una serie de métodos específicos con análisis detallados de cada uno —los dominios de su utilidad, la justificación de sus acciones, etcétera. En el título que sirve de etiqueta a esta segunda concepción, nos hemos referido a la inteligencia artificial como a “la ciencia de los métodos débiles”. Ahora hay que explicar por qué.

Un método tiene dos aspectos. Por un lado, exige cierta información sobre cualquier tipo de tarea a la que se pretenda aplicarlo. Esto se manifiesta en los “datos” del enunciado del problema. Tal exigencia de información puede ser rígida o flexible.

Por otro lado, el método produce ciertas cosas en forma de resultados —o al menos, posibles resultados— con distinto gasto de esfuerzo. Además, puede producir mucho o poco, y puede hacerlo con certidumbre o con verosimilitud, a bajo precio o con un gran coste. En general, cuanta mayor información se pueda utilizar, mejores serán los resultados a obtener. Y a la inversa, los

resultados de mayor interés implicarán una exigencia más fuerte de información, y las exigencias débiles sólo pueden acarrear resultados débiles.

Si comparamos los métodos de la inteligencia artificial con los del análisis numérico, por ejemplo, o con los de la programación matemática, nos llama la atención no sólo el hecho de que los primeros producen resultados débiles, sino que además exigen poca información. Consideremos el más débil de todos ellos, Generar y Comprobar. Es necesario conocer un generador del conjunto y un comprobador de elementos, pero nada más. Por ejemplo, el espacio no tiene que ser numérico ni tener ninguna otra propiedad especial. Cuando nos trasladamos a la Búsqueda Heurística, el enunciado del problema exige algo más: que los nuevos puntos del espacio se puedan obtener a partir de puntos antiguos por medio de operadores. En efecto, hemos especificado una estructura particular para un generador de elementos en el espacio. El método de Búsqueda Heurística es, simplemente, la organización adecuada para potenciar ese nuevo conocimiento, manteniendo en memoria los estados ya generados y regresando a los viejos elementos para iniciar la búsqueda en otras direcciones.

La Ascensión Máxima añade aún otro requisito: que seamos capaces de comparar la proximidad de elementos distintos al elemento deseado (lo que no se necesitaba ni en Generar y Comprobar ni en Búsqueda Heurística). En consecuencia, el dispositivo de retroceso puede ser eliminado del procedimiento en el caso de que el espacio sea unimodal. Pero para proceder al método de la ecuación de segundo grado (Cuadro 1.4), se necesita una especificidad mucho mayor: una ecuación, una variable única cuyo dominio son los números reales, y un polinomio de segundo grado. El procedimiento es, de acuerdo con ello, preciso.

En resumen, esta concepción de la inteligencia artificial no quiere decir solamente que la inteligencia artificial posea métodos, sino que éstos, los métodos de la inteligencia artificial, ocupan un lugar particular en el conjunto de todos los métodos; es decir, son aquellos métodos que requieren muy poca información y que, sin embargo, ofrecen alguna posibilidad de obtener resulta-

dos. Estas condiciones son exactamente las que se dan en situaciones problemáticas.

TERCERA CONCEPCION:

PSICOLOGIA TEORICA

Los modelos del hombre se construyen de la misma forma que se construyen los modelos de cualquier otra cosa: proponiendo sistemas simbólicos susceptibles de ser identificados con hombres específicos en situaciones específicas por medio de la observación y el experimento. El grado de formalismo de tales modelos* puede variar ampliamente, desde conjuntos de afirmaciones en lenguaje natural que dejan casi todas las cuestiones abiertas (por ejemplo, el modelo de Freud sobre la mente tripartita), hasta modelos totalmente formalizados que son dechados de pureza científica, aun cuando resulten un tanto limitados (por ejemplo, algunas teorías estocásticas del aprendizaje). Los problemas de todos los modelos son los mismos, y se derivan de su papel general en el desarrollo de la ciencia: alcance, adecuación a los datos básicos seleccionados, capacidad de ampliación, potencia predictiva, comunicabilidad, simplicidad, etcétera.

La estructura teórica de la ciencia no se reduce a un conjunto de entidades aisladas llamadas teorías. Las teorías guardan entre sí recíprocas relaciones de deducción, y son inferidas y construidas para hacer frente a un cuerpo de datos. Dentro de nuestra inveterada tradición escolar de hacer referencia a nuestros progenitores, es usualmente posible rastrear estas relaciones genéticas. Pero las teorías muestran también entre sí un parecido de familia por otras muchas razones. En todo momento y para todo campo de fenómenos, existe una estructura subyacente de ideas sobre los ingredientes de una teoría. Si se desea construir una teoría de X, entonces, además de darse la posibilidad de modificar la teoría ya existente de X, se sabe qué tipos de lenguajes

* Utilizo indiferentemente los términos *modelo* y *teoría*.

teóricos pueden ser útiles, qué tipos de leyes pueden funcionar con éxito, qué tipos de técnicas de derivación pueden aplicarse (configurando así los ingredientes para permitir que actúen tales técnicas), y así sucesivamente. Todo este conocimiento opera en el contexto de la construcción (es decir, el descubrimiento), no en el contexto de la verificación de teorías. Gran parte de ello representa un entramado que será sustituido cuando la teoría esté completa.

Existen muy pocas descripciones pormenorizadas de esta estructura del conocimiento, bien sea que se la utilice para una teoría particular o bien para un campo total. Por lo demás, ni siquiera existe todavía un vocabulario adecuado con el que describir tales cosas. A menudo hablamos del *Zeitgeist* [espíritu del tiempo] y, desde Kuhn (1962), de *paradigmas*, que es algo más preciso. Pero es digno de notar que los *Zeitgeists* y los paradigmas nunca son “exhibidos”; por el contrario, se los designa señalando las teorías que de ellos se derivan (por ejemplo, el “paradigma de la física clásica”), o emitiendo unos cuantos enunciados sobre algunos de sus ingredientes.

Un aspecto de esta amplia estructura lo constituyen las clases de sistemas en cuyo seno se indaga una teoría particular. El mejor ejemplo, sin duda, lo proporciona la física, donde la teoría de ecuaciones diferenciales (y hasta cierto punto de integrales) ha jugado este papel. La mecánica de partículas, la corriente térmica, la cuerda vibratoria, la elasticidad, la hidrodinámica, el flujo plástico, la electrodinámica —todas ellas, en algún momento de su desarrollo, fueron teorías distintas, y todas utilizaron el mismo esquema básico: un conjunto de números reales para denotar un estado, y un conjunto de ecuaciones diferenciales para representar los cambios de ese estado.

En muchos campos cabe utilizar distintos lenguajes. En neurofisiología, por ejemplo, se puede pensar en términos de redes de neuronas, jugando un importante papel la conexión y la interacción neurona-neurona. O se puede pensar en términos de campos de variación continua (como en el EEG), jugando un importante papel los conceptos de correlación y relaciones de fase.

Y, recientemente, se han añadido los esquemas moleculares y bioquímicos (por ejemplo, en la investigación de la memoria).

La psicología tiene al menos seis lenguajes. Uno es la teoría $S - R$, representada por distintos conjuntos de estímulos y respuestas que varían en su fuerza de conexión y en las probabilidades de respuesta. Un segundo lenguaje está constituido por las teorías estocásticas del aprendizaje, donde el estado del organismo está representado por medio de las probabilidades de respuesta, con operadores sobre esas probabilidades en forma de leyes. Un tercero serían los modelos de Markov, que consideran al organismo como estando en un determinado estado discreto, con leyes que se establecen en forma de una matriz de transición de probabilidades entre estados. Un cuarto sería el que considera a la persona como un conjunto de “fuerzas”, que presionan de una manera u otra, como en el campo vital de Lewin. Un quinto sería la persona como una comunidad interactuante de subagentes, como el Id, el Ego, y el Super-Ego de Freud, que son susceptibles de ser descritos en términos de personalidad. Un sexto, que tiene aquí para nosotros especial interés, es el hombre como un procesador de información. Y todavía existen más —por ejemplo, los modelos de sistemas de control, utilizados en estudios del operador humano, no derivan de ninguno de los anteriores, y lo mismo se puede decir de los modelos de detección de señales.

Una característica importante de tales lenguajes, o clases de sistemas, es que ellos mismos no son teorías. No son correctos o incorrectos, en el sentido de que se puedan probar mediante una prueba empírica. Son generadores de teorías, todas las cuales tienen una semejanza de familia y se supone que comparten aciertos y debilidades. Estos lenguajes serán o no útiles en la medida en que las teorías particulares que producen sean buenas o malas (o, si se desea, se puede describir la calidad de las teorías enfrentándolas con los fenómenos que pretenden explicar).

Esta digresión sobre un aspecto general de las teorías científicas es útil para poner los cimientos de la tercera concepción de la inteligencia artificial: que sirve como psicología teórica si se adopta una concepción del hombre como procesador de informa-

ción, representado por un sistema de símbolos discretos. En consecuencia, el lenguaje utilizado para construir teorías particulares es el que se ha desarrollado en la programación de computadores, y los tipos generales de sistemas que se diseñan son los desarrollados en inteligencia artificial. (Como más adelante veremos, es necesario ampliar esta cuestión).

El dominio de tales teorías no abarca todos los fenómenos que rodean al hombre. Como cabría esperar, se centra principalmente en tareas que son en alto grado simbólicas y que no necesitan continua habilidad motora o una íntima dependencia de los sistemas sensoriales. Se enfrenta con fenómenos de complejidad sustancial, pero no con la total complejidad de los asuntos humanos.

La concepción del hombre como procesador de información está actualmente extendida en muchas partes de la psicología. Esta concepción tiene también raíces y fuentes de inspiración que van más allá del ámbito de la inteligencia artificial, tal como la hemos descrito en este trabajo. Según los criterios que se adopten para enumerarlos, existen de tres a seis subcampos: solución de problemas, psicolingüística, formación de conceptos, aprendizaje verbal, memoria inmediata y percepción.* Los cuatro últimos se pueden agrupar de distintas maneras. En particular, el trabajo sobre percepción se ha desarrollado a partir de la investigación sobre memoria inmediata, y constituye así un aspecto limitado (aunque importante) de todo lo que aparece bajo la rúbrica de percepción. El trabajo sobre el desarrollo acaba de comenzar actualmente (DIGGORY, 1972).

La diversidad de fenómenos y fuentes comporta hasta cierto punto una diferenciación del producto, pero no oscurece la

* Los aspectos de la ciencia de las organizaciones humanas y de la microeconomía (la teoría conductual de la empresa), se deberían incluir en cualquier enumeración completa. El caso de la psicología del desarrollo es curioso, puesto que la obra de Piaget, que está teniendo un impacto abrumador, es, por un lado, bastante similar al procesamiento de información (por ejemplo, en su énfasis en las operaciones), y, por otro lado, es muy distinta (por ejemplo, en su total ausencia de interés por el procesamiento de representaciones).

identidad fundamental de las concepciones de procesamiento de información que operan en todos estos dominios. Aunque nosotros diferenciamos las diversas regiones en alguna medida, no es posible hacerlo por completo, puesto que los científicos que trabajan en este campo leen la literatura de los demás especialistas que también lo cultivan y todos ellos están mutuamente al tanto de las investigaciones ajenas.

La concepción de que el hombre es un procesador de información significa que su conducta puede ser considerada como el resultado de un sistema compuesto de memorias que contienen símbolos discretos y expresiones simbólicas (o sea, ocurrencias de símbolos), y procesos que manipulan estos símbolos. El concepto central es el de símbolo, que significa esencialmente lo mismo que en las ciencias de la computación, una entidad con cierta propiedad funcional, o sea: que cuando un proceso se enfrenta con una instancia de un símbolo, tiene acceso a información (codificada en expresiones simbólicas) sobre lo que ese símbolo designa. Los procesos que pueden realizarse con símbolos son su creación (y, posiblemente, su destrucción), la obtención de la información designada, la creación de expresiones simbólicas, y la manipulación de estas expresiones simbólicas mediante inserción, supresión, sustitución y reordenación. Una expresión simbólica es un conjunto de instancias concretas de símbolos conectados por relaciones de acceso —por ejemplo, las instancias o marcas S, I, M, B, O, L, O conectadas por la relación *siguiente*, que proporciona una señal de la palabra SIMBOLO.

Se puede intentar formalizar estas nociones más detalladamente, es decir, crear una definición única para una clase de sistemas de tal forma que todo aquello que sea un sistema simbólico discreto, sea un ejemplo de esa clase. Realizar esta tarea importa menos que comprender los tipos de sistemas implicados. Lo que cuenta son los sistemas particulares producidos dentro de esta tradición. Cada uno de ellos se sostendrá por sí mismo en la medida en que afronte las pruebas de verificación. El tipo de descripción a que hemos recurrido (y sus ampliaciones) basta para dar cuenta de la semejanza familiar de estos modelos.

De acuerdo con los tipos de procesamiento que permiten realizar, este modelo general constituye precisamente, cuando se lo formaliza, otra variante de los procedimientos efectivos, máquinas de Turing, etc. Una vez más, el interés no estriba en esto, sino en lo satisfactoriamente que traten la conducta humana los modelos de este tipo. Consideremos algunos ejemplos.

Extrapolación de secuencia

Puesto que ya hemos considerado el modelo de Simon-Kotovsky para manipular secuencias extrapoladas, también podemos observar su papel como modelo de conducta humana. No existe mucha dificultad en realizar una identificación de la teoría con la conducta humana, puesto que se le puede pedir al programa que realice predicciones de secuencias, de la misma manera que se le puede pedir a un hombre. La primera comparación evidente es que tanto el modelo como el hombre solucionan problemas, de forma que el modelo proporciona una explicación de cómo es posible que un hombre solucione estos problemas. Con frecuencia, se hace referencia a esto como a una *prueba de suficiencia*, que es muy característica de los modelos de procesamiento de información, en especial contraste con la mayoría de las demás teorías psicológicas, que no son lo bastante comprehensivas como para realizar tales propósitos, sino que son, por el contrario, más analíticas (que aquí entendemos en el sentido de opuesto a sintético), al tratar de relaciones entre aspectos abstractos de la conducta. En consecuencia, para tareas complejas, tales como las tareas de inducción, simplemente no existen teorías acerca de cómo realizan los sujetos humanos esas tareas salvo las teorías pertenecientes a la familia del procesamiento de información. Dentro de ello, naturalmente, se pueden generar posibles alternativas al programa de Simon-Kotovsky.

Nosotros deseamos algo más que la suficiencia. Por ejemplo, podemos desear comparar la dificultad de las tareas para el modelo y para el sujeto humano. Esto se realizó en el estudio ori-

ginal (SIMON-KOTOVSKY, 1963), utilizando el éxito/fracaso como medida de dificultad, con una concordancia bastante razonable en algunas variantes del modelo y menos deseable en otras. Los autores de ese estudio llegaron algo más lejos al poner en relación la dificultad con características estructurales de las descripciones de las series, como, por ejemplo, el número de variables indicadoras que había que mantener: cuantos más indicadores, mayor era la dificultad de la tarea. Esto está bastante de acuerdo con otros datos sobre la importancia de las limitaciones de la memoria inmediata.

El modelo total (el programa) conlleva una serie de supuestos, y para comprobarlos y verificarlos todos se necesita más evidencia conductual que la proporcionada por las medidas de realización. El interés por tomar un contacto masivo con la cantidad de conducta, como ayuda para la identificación de teorías, es también característico del enfoque orientado al procesamiento de información. Consideramos a continuación un ejemplo demostrativo de este punto.

Solución de problemas en criptoaritmética

He trabajado ampliamente en problemas de criptoaritmética, donde hay que asignar distintos dígitos a letras, de forma que la suma resultante sea correcta (NEWELL y SIMON, 1972). (El Cuadro 1.11 proporciona un ejemplo donde CROSS + ROADS es igual a DANGER). Su virtud como tarea de solución-de-problemas (al margen del hecho de que son divertidos) descansa en la combinación de las actividades de búsqueda, ensayo y razonamiento que ponen de manifiesto la mayoría de las personas de educación superior cuando realizan la tarea.

Los datos de conducta que se obtienen en tales tareas son las verbalizaciones del sujeto mientras está trabajando (además de los registros escritos, en el caso de que se permitan). Tales datos permiten inferencias no sólo sobre el resultado final (como en el trabajo de Simon-Kotovsky), sino sobre lo que el sujeto conoce

en muchos puntos intermedios durante la solución del problema. De esta forma, en la tarea de $CROSS + ROADS = DANGER$, la interpretación de la expresión "D debe ser 1", como significativa de que el sujeto sabe ahora que D debe ser 1, supone ya un sistema capaz de "tener conocimiento" y de tratar con "símbolos" que "tienen significado". Estos supuestos son esencialmente los inherentes a un sistema de procesamiento de información. Así, parece poco peligroso aceptarlos. Dicho de otra forma, las teorías de procesamiento de información proporcionan un modelo de significado que justifica la interpretación asignada a expresiones tales como la anterior.

Más importante es el tipo de inferencias que pueden realizarse a partir de las verbalizaciones (o de los escritos correspondientes).^{*} Si el sujeto dice "Sea S igual a 1", ¿es que no sabe, por tanto, que D debe ser 1? Probablemente es una buena hipótesis de trabajo. Si después el sujeto dice, "Pero D es lo que nos llevábamos al escribir la última columna a la izquierda, de modo que debe ser 1; por tanto, S no puede ser 1", entonces la hipótesis se puede tomar como un hecho. Es decir, puede utilizarse para apoyar o para desconfirmar determinados modelos que realizan predicciones sobre lo que el sujeto conoce acerca de S y D. Naturalmente, es posible pasar de expresiones tan simples como las anteriores a párrafos de una oscuridad casi arbitraria; "Ahora si la — ¡oh!, lo siento, he dicho algo equivocado. Estoy haciendo —no, no, ninguno de los dos". Una transcripción cargada de tales pasajes representa simplemente una serie de datos que no se pueden analizar.

Para iniciar y validar un modelo de procesamiento de información, nos gustaría saber el estado completo del conocimiento del sujeto en cada momento, limitándose entonces el modelo a proporcionar una representación de este conocimiento y de las transformaciones que se realizan. El conocimiento es análogo a la

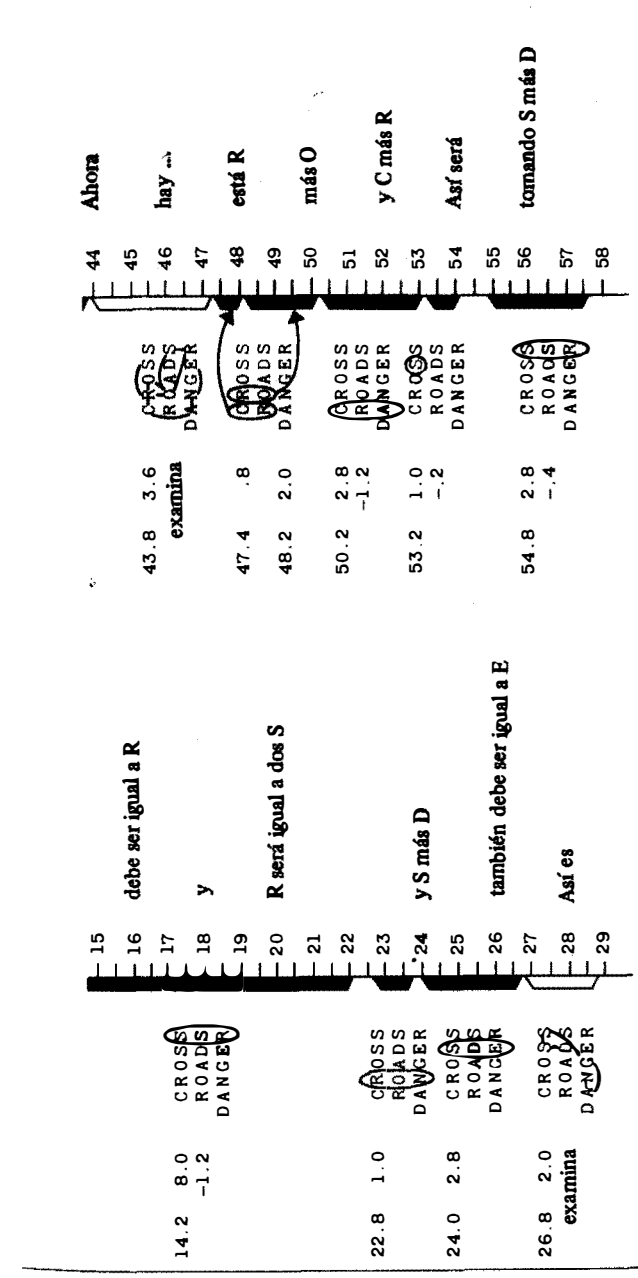
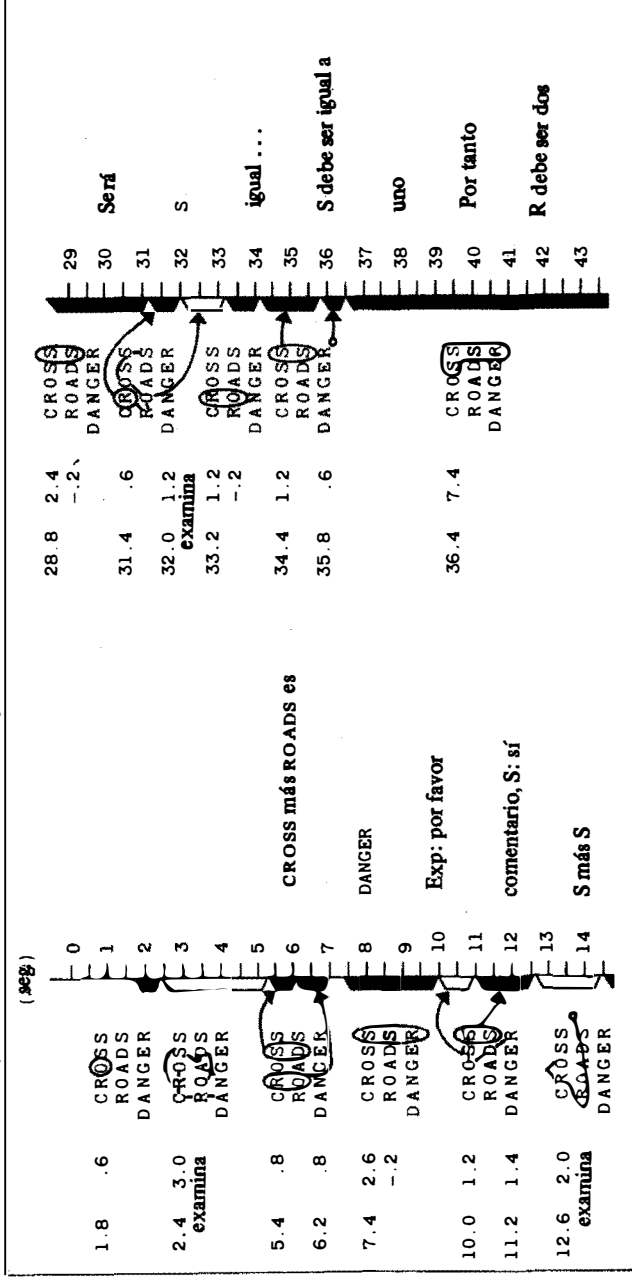
^{*} La discusión implica que la inferencia que va de las verbalizaciones al estado de conocimiento del sujeto, es realizada por un humano. Trabajos recientes indican que este proceso puede ser mecanizable (WATERMAN y NEWELL, 1971).

energía en el hecho de que su cantidad absoluta no puede ser determinada, sino sólo los cambios que en su seno ocurren. Por tanto, en una tarea como la de criptoaritmética, una gran parte de lo que el sujeto conoce no cambia durante la solución del problema, ni forma parte de la selección de las transformaciones. Todo lo que se debe saber sobre el estado del conocimiento es la parte relevante. Aun así, naturalmente, los datos de que se dispone permiten una inferencia directa (esto es, sin el beneficio del modelo específico a construir) de solamente una fracción de lo que se desea.

Podríamos intentar obtener más datos simultáneamente, durante la solución del problema. Tales datos deberían confirmar en parte (o dejar de confirmar) los hechos sobre el estado del conocimiento ya derivado de las verbalizaciones, pero también podrían añadir nuevos datos. En última instancia, deseamos tener muchas medidas distintas, incluyendo las fisiológicas, aunque estas últimas aún parecen estar muy lejanas. Los datos sobre movimientos oculares son una posibilidad que se ha explorado y parece que proporcionan una información adicional (WINIKOFF, 1967). El Cuadro 1.11 muestra tanto las verbalizaciones como los movimientos oculares, durante unos cuantos minutos, de un sujeto que realiza la tarea de $CROSS + ROADS = DANGER$. El corto desarrollo de la figura no muestra fijaciones, sino más bien agregados de fijaciones en unidades llamadas *unidades de atención* y *unidades de examen*.

Estas unidades tienen la importancia funcional indicada por sus nombres. Durante una unidad de atención se debe considerar que el sujeto está atendiendo a una parte específica de lo que se exhibe, y durante una unidad de examen está buscando algún tipo de información. Lo cual quiere decir: en cualquier modelo de la conducta del sujeto debemos tomar estos movimientos oculares, así interpretados, como confirmando (o desconfirmando) el modelo propuesto. Una gran parte del Cuadro 1.11 no está explicada aquí, como por ejemplo la derivación de las unidades agregadas. También existen *unidades de transición*, que son de breve duración y pueden ocurrir entre unidades de atención (ob-

Cuadro 1.11 Conjuntos de Movimientos Oculares y Conducta Verbal



sérvense las pequeñas lagunas en la columna vertical de tiempo), aunque también dentro de las unidades de atención. El trabajo de Winikoff contiene una discusión exhaustiva de todos estos aspectos, incluyendo la validez de la correspondencia entre las unidades de movimiento ocular y la conducta verbal.

Lo que se desea es crear un modelo de procesamiento de información que se conduzca como el humano que soluciona problemas, concebido a través de las inferencias realizadas a partir de los datos verbales y de movimientos oculares. Naturalmente, estos datos son muy específicos: un episodio por cada sujeto. Últimamente el interés se centra en la generalización sobre las tareas y sobre las personas, como también en averiguar cómo encaja este sistema de solución de problemas (suponiendo que lo tengamos) en la estructura del sistema total de procesamiento de información del sujeto. De todas formas, la tarea así planteada es razonable (casi paradigmática) para una ciencia de la conducta humana.

Para encontrar un modelo construimos —como se muestra en la Figura 1.2— una representación gráfica de las variaciones en el estado de conocimiento del sujeto tal como se infieren de los datos, a lo que se llama Gráfico de Conducta de Problema (GCP). Puede entenderse como una búsqueda a través de un espacio de posibles estados de conocimiento sobre el problema, llamado *espacio del problema*. Los medios de búsqueda son un conjunto de operadores: en el GCP de la Figura 1.2 aparecen cinco de tales operadores. PC elabora una columna determinada del planteamiento, realizando cualquier inferencia que esté a su alcance (por ejemplo, si $S = 1$, el PC (columna 1) $\Rightarrow R = 2$). AV realiza una asignación tentativa de un dígito a una variable, ya sea una letra o un arrastre de la suma (por ejemplo, sea S igual a 1). TD comprueba si un dígito puede ser legítimamente asignado a una letra (por ejemplo, si $D = 1$ entonces TD concluirá que $S = 1$ no es posible). FC encuentra una columna que contiene una letra o un arrastre de suma determinado (por ejemplo, $FC(R) \Rightarrow$ columnas 1, 4, 5). FNC encuentra la siguiente columna por el método típico para realizar la suma (por ejemplo, si la columna 1 acaba de ser sumada, entonces $FNC \Rightarrow$ columna 2). Dos operadores de memo-

ria no aparecen en GCP: RA, que recuerda la situación antecedente; y RV que recuerda el valor de una variable.

Cada uno de estos operadores, cuando se ejecuta con éxito, proporciona nuevo conocimiento (incluyendo información atencional —dónde buscar a continuación). Esto nos lleva a que el sujeto puede ser descrito como algo que obtiene todo su conocimiento de la aplicación de este pequeño conjunto de operadores. El conjunto de posibles estados de conocimiento dentro de los que se desplaza el sujeto, puede describirse mediante una gramática BNF, de la misma forma que el espacio de las hipótesis para el programa de inducción.

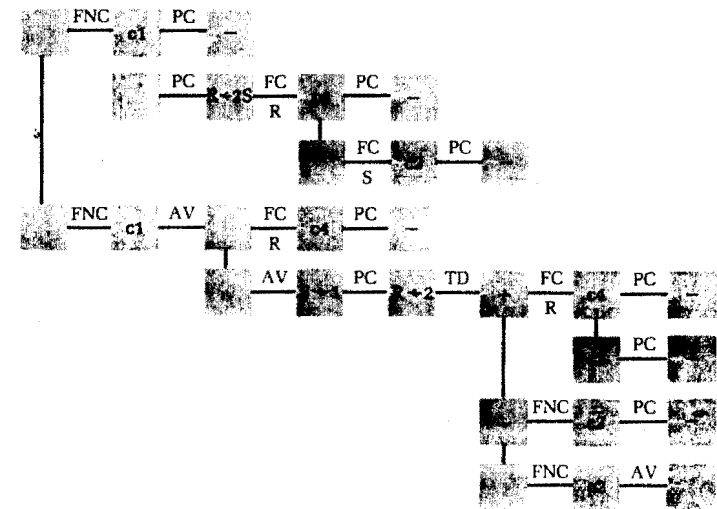


Figura 1.2 Gráfico de Conducta del Problema (GPC).

El árbol de búsqueda de la Figura 1.2 proporciona una forma de estructurar un modelo de procesamiento. En cada nódulo, existe un cierto estado de conocimiento y existe la selección de algún operador para avanzar en el conocimiento. Algunas veces, naturalmente, lo que se selecciona es la acción de regresar a un lugar anterior en el árbol de búsqueda para comenzar una nueva en otra dirección. En cada punto, podemos describir una regla que informa de la selección como una función de la información en el estado actual del conocimiento. No se necesita una regla distinta para cada nódulo; la misma regla puede aplicarse a muchos nódulos, dando por resultado una variación en la conducta de diferencias en el estado del conocimiento, en lugar de diferencias en las reglas. La totalidad de tales reglas forma un programa de acción que es análogo en muchos aspectos a un algoritmo normal de Markov o a un sistema canónico de Post. Se ha convertido en algo común hacer referencia a cada regla parcial como a una *producción*, constituyendo la totalidad un *sistema de producción*.

El Cuadro 1.12 muestra un sistema tal de producción para la conducta manifestada en el Cuadro 1.11. Las reglas tienen una parte de condición y una parte de acción. En la estructura de control, cada producción se considera por turno. Si se satisface su condición para el estado actual de la memoria inmediata, entonces se llama a su acción, ocurre algún cambio en la memoria inmediata, y el sistema comienza de nuevo en el principio. Si no se satisface ninguna condición, el procesamiento termina (a menos que aparezca nueva información desde el exterior, por ejemplo mediante percepción de una configuración externa). La memoria inmediata está compuesta de un aglomerado de metas al igual que de un conjunto de afirmaciones elementales sobre letras y dígitos. Lo que no está representado es el conocimiento de los caminos a seguir ni los estados del conocimiento a los que es posible regresar.

Consideremos la primera producción, S1. La condición es que una expresión de la forma "letra = nuevo dígito" ocurra en la memoria inmediata. *Letra* y *dígito* son variables de los tipos indicados; "=" es la relación específica de igualdad, y "nuevo" es uno de los dos calificadores específicos, siendo el otro $\square$, el signo de

Cuadro 1.12 Sistema de Producción y Estado de Conocimiento

S1: letra = nuevo dígito $\rightarrow$ FC(letra) $\Rightarrow$ columna; usar columna
 S1': letra $\leftarrow$ nuevo dígito $\rightarrow$ hacer PC
 S1'': usar columna $\rightarrow$ PC
 S2: obtener variable $\rightarrow$ FC(variable) $\Rightarrow$ columna; PC(variable, columna)
 S3: usar columna, PC(columna) $\Rightarrow$ $\rightarrow$ AV(letra(superior), columna)
 S4: letra desigualdad dígito $\rightarrow$ hacer AV(letra)
 G2: letra $\leftarrow$ dígito $\square$ $\rightarrow$ hacer AV(letra)
 G4: ϕ | + | - $\rightarrow$ FNC $\Rightarrow$ usar columna
 T1: letra = nuevo objeto, $\neg$ hacer TD $\rightarrow$ hacer TD(letra, objeto)
 T2: letra = dígito $\square$ $\rightarrow$ RA $\Rightarrow$ expresión; expresión $\square$
 R2: comprobar variable $\rightarrow$ RA(variable) $\Rightarrow$ producción; repetir producción
 R3: evocar letra $\rightarrow$ RA(letra) $\Rightarrow$ columna; RV(letra, columna)

variable := letra | arrastre

objeto := dígito | letra

expresión := variable relación objeto calificador

relación := $\leftarrow$ | = | desigualdad

desigualdad := $\geq$ | $\leq$ | $>$ | $<$

calificador := $\square$ | nuevo | nulo

operador := PC | AV | TD | FNC | FC | RA | RV

principal hallar evocar

tipo de meta := obtener variable | realizar operador principal | comprobar variable | usar columna | evocar letra

Operación de apilamiento de metas:

señal	interpretación	acción
nulo	nueva meta	utilizarla para seleccionar la sig. producción
+	éxito	extraer otra meta si hay más en la pila
-	fracaso	extraer otra meta si hay más en la pila
i	interrupción	continuar
i-	fracaso condicional	si la submeta se logra continuar; en caso contrario fracaso

Acciones de producción: si hay un tipo de meta, entonces introducir la meta actual en la pila; en caso contrario añadir la información producida a la meta actual.

imposible. Así, "R = nuevo 1" sería consistente con la condición, pero "R = 1" o "S > nuevo 1" no sería consistente. La condición de S1 consiste en encontrar una columna que contenga la letra (FC) e introducir en la memoria inmediata la expresión "usar columna". La columna llega como una salida de FC, que se indica convencionalmente mediante "⇒ columna". "Usar" es una de las diversas metas, siendo las otras "colocar", "hacer", "comprobar", "recordar", y "repetir". Las metas se introducen en el conjunto de metas de la memoria inmediata y son automáticamente pre-sionadas y propulsadas. Así, la distinción entre las acciones "PC" y "hacer PC" es que la última, pero no la primera, se introduce en el conjunto de metas.

Esta descripción, aunque esquemática, quizá es suficiente para permitir una apreciación de cómo actúa el sistema. El Cuadro 1.13 muestra la conducta del sistema en unos pocos pasos. Nosotros habíamos comenzado con una memoria inmediata que corresponde a $t = 14.2$ en el Cuadro 1.11 y la Figura 1.2, y la recorre hasta $t = 43.8$. Esto sería suficiente para mostrar la naturaleza de la correspondencia —que el programa consiste en predecir qué operadores deben ser realizados en cada paso, por tanto qué estado de conocimiento se debe alcanzar en cada paso, y de esta forma qué cosas relevantes para la tarea pueden manifestar los sujetos en cada punto, ya sea mediante verbalizaciones o por movimientos oculares. No existe nada en el modelo que determine qué se dirá realmente o qué es posible decir. La comparación con los datos consiste en que cada dato correcto cuenta *a favor*, cada dato incorrecto cuenta *en contra* y se ignoran las omisiones.

Sigue siendo cierto que la comparación más informativa es la que se realiza directamente entre la salida del sistema y la conducta registrada. Además, podemos resumir la bondad de un modelo contabilizando cuántas producciones se han utilizado para abarcar una cantidad determinada de datos, donde estos últimos se miden en términos del número de nódulos en el Gráfico de Conducta de Problema. Tal gráfico se muestra en la Figura 1.3, donde hemos desarrollado las diversas producciones en el orden del número de nódulos que cubre cada uno. Esto pone en eviden-

Cuadro 1.13 Huella del Sistema de Producción

Tiempo	Producción	Operador	Apilamiento de meta (PC cl i)	Expresiones
14.2	continue	PC	(get P)(get S) (PC cl i-)	(P = 2S new)
22.2	S2	FC		
22.8	↑	PC	(get P c4)(get S) (PC cl i-) (get R c4-)(get S) (PC cl i-)(get S) (PC cl i-)	
23.8	S2	FC	(get S c2)(PC cl i-)	
24.0	↑	PC	(PC S c2-)(PC cl i-) (PC cl i-)	
26.8	G4	FNC	(use cl)	
28.8	S3	AV	(get R)(AV S cl i)	
31.2	S2	FC	(get R c4)(AV S cl i)	
31.4	↑	PC	(PC R c4-)(AV S cl i) (AV S cl i)	
34.4	continue	AV	(AV S cl +)	(S ← 1 new)
36.4	S1'	PC	(PC cl +)	(R = 2 new) (S ← 1)
	T1	TD	(TD R +)	(R = 2 new)
43.8				

Notas: "Continúa" implica que tiene el control el operador del apilamiento de metas.

↑ significa que la producción precedente continúa teniendo el control.

La extracción del apilamiento de metas bajo las señales de meta +, - o i- ocurre automáticamente antes de que se seleccione otra producción.

Los tiempos son resultado de la comparación de los conjuntos de movimientos oculares con la acción que está siendo realizada por el operador.

cia que algunas producciones juegan un papel fuerte y otras uno mucho más débil —y, al mismo tiempo, que algunas reciben mucho soporte de los datos y otras poco. Una producción ha de ocurrir al menos dos veces para evitar que sea completamente *ad hoc*.

El sistema del Cuadro 1.12 debe verse como una descripción de procesamiento de información de la conducta del sujeto, puesto que fue diseñado para describir estos datos. La fuerza de esta descripción es doble. Por una parte, descansa en su relación racional con la tarea que se realiza. No es, por tanto, un sistema arbitrario (como, por ejemplo, un polinomio de grado arbitrario, que tiene que ajustarse a un conjunto de puntos) en relación con el sistema que describe. Ni tampoco es único, puesto que existen otros sistemas que pueden solucionar la tarea (o intentar solucionarla). Por ejemplo, el sistema REF—ARF la soluciona (recuérdense los Cuadros 1.2 y 1.3). Sin embargo, todos estos sistemas tienen un parecido de familia —esto es, son sistemas de procesamiento de información. El segundo aspecto de la fuerza o potencia descriptiva es el que ofrece la Figura 1.3, que es análogo a una medida de ajuste —la cantidad de libertad requerida para describir una serie de datos.

El sistema de producción describe sólo la conducta de este sujeto aislado en este único experimento (de hecho, sólo los primeros 200 segundos). Sin embargo, la conducta de otros muchos sujetos muestra la misma estructura, y aun algunos con las mismas producciones, cuando el análisis se ha realizado con el mismo detalle (NEWELL y SIMON, 1972). Los sujetos no necesitan conducirse de esta forma. Por ejemplo, a algunos sujetos que han intentado construir un conjunto de ecuaciones algebraicas y manipularlas, no se los podría describir completamente mediante este modelo específico. Otra cuestión es que fuese posible describirlos por su trabajo en un espacio de problema, donde los estados de conocimiento incluyen conjuntos de ecuaciones y los operadores implican la manipulación de estas ecuaciones.

Este ejemplo puede considerarse representativo de la labor realizada en solución de problemas utilizando modelos de proce-

samiento de información —la construcción de un sistema explícito de procesamiento de cierta complejidad que realiza la tarea, el uso de datos verbales (y también de datos de movimientos oculares), el intento de ajustar la configuración dinámica de la conducta a la conducta del sujeto. Un trabajo similar, aunque sin incluir sistemas de producción, se viene realizando en tareas de lógica elemental, en ajedrez, y en otras distintas áreas (véase, NEWELL y SIMON, 1972).

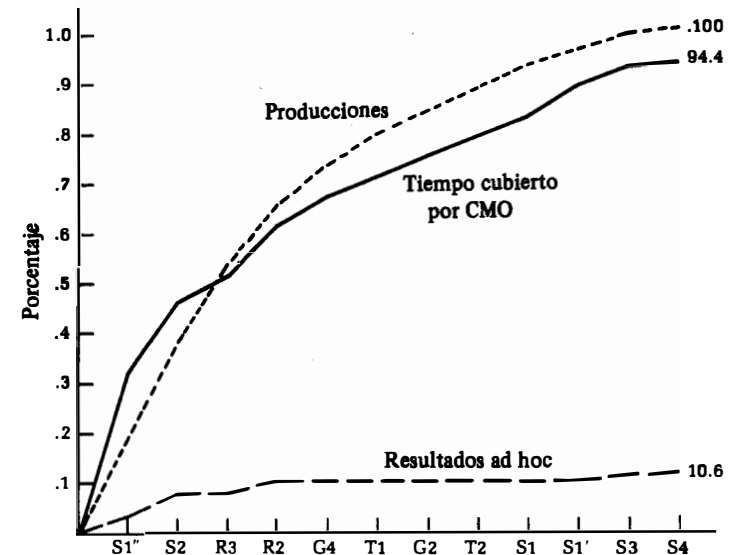


Figura 1.3 Sumario de la Amplitud de Producción.

Memoria inmediata

Aunque he descrito en primer lugar la solución de problemas, la mayor parte del trabajo basado en la idea de procesamiento de información discurre en otras áreas de la psicología. La investigación sobre memoria inmediata nos proporciona un importante ejemplo. Sus raíces se remontan a la ingeniería del control y la comunicación. Donde mejor puede verse esto es en el libro de BROADBENT sobre *Perception and Communication* (1958), en el cual habla su autor sobre la memoria en términos de bucles reverberantes y sistemas de control, motivado por una serie de experimentos sobre percepción auditiva.

Se da también una conexión con la teoría de la información, aunque de tipo negativo. La fascinación en psicología por la teoría de la información tuvo lugar a principios de los años 50, centrándose la atención sobre los conceptos de capacidad y el conjunto sobre el que se realizaba una selección. Los primeros resultados positivos versaron sobre la capacidad total del canal en un hombre que funcionaba como transductor, y manifestaban que el tiempo de reacción era lineal en relación con el contenido de la información del conjunto total. El justificadamente famoso artículo de George MILLER sobre "siete más o menos dos" (1956) proporcionó el marco a este tipo de trabajos, puesto que ofreció evidencia suficiente de que existía un sistema de memoria inmediata, cuya capacidad debía medirse en *chunks* [trozos], no en *bits* [dígitos binarios]. Estos "trozos" son de hecho símbolos.

Desde entonces, el problema de la memoria inmediata ha venido discutiéndose más y más en términos de procesamiento de información. Sin embargo, el estilo tanto de los experimentos como de las teorías es notablemente distinto del trabajo sobre solución de problemas. En este caso se está intentando obtener una información detallada de la estructura de un órgano determinado. Se diseñan modelos detallados de su estructura y procesamiento que permitan un tratamiento matemático, y es muy acusado el gusto por la ingeniería de control y el diseño lógico de estos modelos. En correspondencia con ello, se realizan experimentos pa-

ramétricos repetidos que permitan tanto la estimación de los parámetros del modelo como la evaluación del grado de ajuste. Actualmente existe una considerable cantidad de modelos de memoria inmediata (WAUGH y NORMAN, 1965; BOWER, 1967; ATKINSON y SHRIFFRIN, 1968; NORMAN y RUMELHART, 1972; LAUGHERY, 1969). Todos ellos, salvo los más recientes, aparecen resumidos en un reciente libro de NORMAN (1969) que trata de teorías de procesamiento de información sobre la memoria.

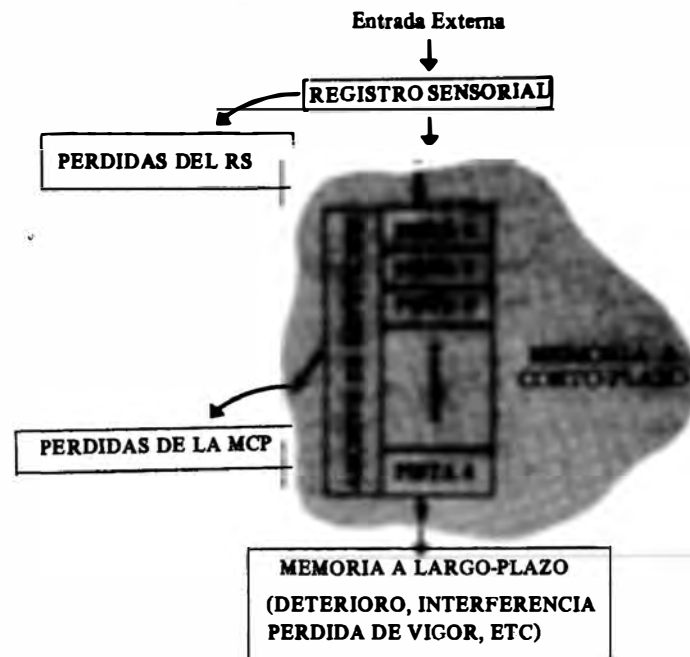


Figura 1.4 Diagrama general de la Memoria Humana (De Atkinson y Shiffrin, 1968).

Para ofrecer un indicio del referido estilo, la Figura 1.4 muestra el diagrama de la estructura de memoria propuesto por Atkinson y Shiffrin. El modelo se centra en el problema de la repetición, e incluye tanto una memoria inmediata como una memoria a largo plazo. Los símbolos regresan a la memoria inmediata bajo el control de una estrategia de ejecución, mantenida por el sujeto en la tarea específica. (Estas estrategias de control no están representadas explícitamente en la teoría, pero tienen la utilidad de justificar en cierta medida sub-modelos específicos de

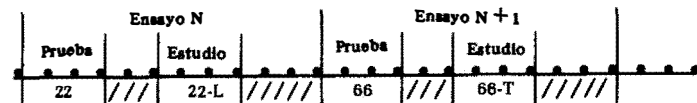
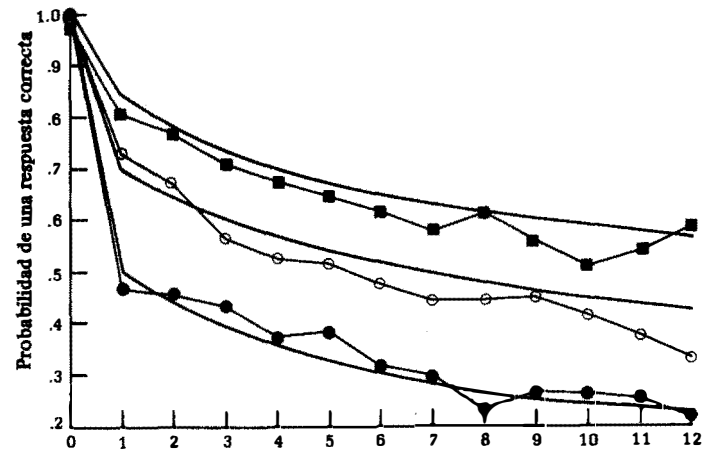


Figura 1.5 Probabilidades Observadas y Teóricas de un Experimento de Memoria (De Atkinson y Shiffrin, 1968).

operaciones de memoria en cada una de las tareas específicas). El olvido en la memoria inmediata se manipula planteando una distribución exponencial de la probabilidad que tendrá un ítem en la memoria. De esta forma, el retorno a la memoria inmediata constituye una repetición, e implica un incremento de memoria de los símbolos repetidos. La transferencia a la memoria a largo plazo se produce como una función del tiempo de permanencia en la memoria inmediata.

La Figura 1.5 muestra algunas curvas teóricas y experimentales de este trabajo. En un ensayo se le preguntó primero al sujeto (PRUEBA) por el valor (una letra) que previamente había estado asociado con un atributo (un número de dos dígitos); a continuación se le fue dando un nuevo valor para asociarlo con el mismo atributo (ESTUDIO). Así, la situación parecería ser, en parte, una tarea típica de pares asociados, pero con una re-asociación continua de nuevos valores. Las curvas exploran los efectos de introducir una cierta cantidad de repetición, siendo la curva superior de tres refuerzos, de dos la curva intermedia, y de uno la curva inferior. La importancia de tales curvas, además de mostrar que el modelo se ajusta a los datos paramétricos, descansa en el análisis detallado del mismo, que no es posible proporcionar aquí. Nuestro propósito consiste en indicar la amplitud del trabajo hoy en marcha donde el sujeto humano se concibe como un sistema de procesamiento de información.

Psicolingüística

Una tercera línea de trabajo que refleja una concepción de procesamiento de información de la conducta humana se deriva de la lingüística, especialmente de la moderna lingüística transformacional. El trabajo de Chomsky y sus colaboradores ha tenido efectos en planos muy distintos de la lingüística. Su impacto sobre la psicología (sobre la psicolingüística) incidió originalmente sobre la realidad de la estructuración gramatical del lenguaje (por ejemplo, el carácter productivo de la conducta lingüística, la reali-

dad de las transformaciones, etc.). Uno de los resultados fue el amplio trabajo sobre la caracterización de la gramática infantil (véase LENNEBERG, 1964). Otro consistió en los experimentos que suponían que las transformaciones, tales como la negación, la pasiva, etc., correspondían a operaciones encubiertas, e intentaban demostrar su realidad —por ejemplo, mostrando la aditividad del tiempo relativo para descifrar las sentencias (véase MILLER, 1962). Recientemente, el impacto se ha extendido a la semántica bajo el impulso final de los propios lingüistas (por ejemplo, KATZ y FODOR, 1963).

Es innecesario trazar un panorama detallado de este trabajo en psicolingüística. Lo importante aquí es que los modelos de lingüística transformacional son una especie de modelos de procesamiento de información, por cuanto incluyen sistemas de símbolos y reglas para operar sobre los símbolos (CHOMSKY, 1965). Existen diferencias, naturalmente. Quizá la más importante se relaciona con la insistencia de los lingüistas sobre la distinción entre competencia y realización. Los modelos formales del lenguaje describen la competencia, impidiendo así pasar a la conducta real de la realización de un locutor o un oyente. Pero el psicolingüista debe encontrar alguna forma de cubrir esta laguna, puesto que lo que desea es explicar realizaciones.

Esto puede ser ilustrado por un estudio reciente de CLARK (1969) sobre problemas de inferencia simple formulados en lenguaje natural. El Cuadro 1.14 muestra dos de estos problemas, tal como fueron presentados a los sujetos. Clark, motivado por consideraciones lingüísticas de estructura profunda, desarrolló tres principios que permiten la deducción del tiempo relativo para responder a tales problemas. De forma resumida, estos principios son:

Primacía de las Relaciones Funcionales. Las relaciones funcionales, como las de sujeto, verbo y objeto directo, están almacenadas inmediatamente después de la comprensión de una forma más fácilmente utilizable que otro tipo de información, como la del tema.

Indicación Léxica. Determinados adjetivos están marcados y sólo pueden existir en forma de contraposición (“¿Es muy malo X?” implica siempre que X es malo); otros adjetivos no están marcados y pueden

existir tanto en forma contrapuesta como en forma nominal (“¿Es muy bueno X?” permite como respuesta “X es bueno” o “X es malo”). El sentido nominal de un adjetivo se almacena y se recupera de la memoria más rápidamente que su sentido contrapuesto.

Congruencia. Se investiga en la memoria una respuesta vinculada información congruente, en el nivel de relaciones funcionales, con la información requerida en la pregunta. La pregunta no puede ser respondida hasta que su reformulación es congruente con la información almacenada (que no se puede utilizar para la reformulación).

Cuadro 1.14 Problemas de Razonamiento Lingüístico (De Clark, 1969)

Si Juan es mejor que Pedro
Y Miguel es peor que Pedro
¿Quién es el mejor?
Pedro Miguel Juan

Si Juan no es tan bueno como Miguel
Y Juan no es tan malo como Pedro
¿Quién es el mejor?
Pedro Miguel Juan

Estos principios tienen dos aspectos. Uno es el uso de conceptos lingüísticos para estructurar la situación: las relaciones funcionales y los adjetivos marcados o no marcados vienen determinados por la evidencia lingüística, que es independiente del experimento. El otro aspecto es el uso de los supuestos de procesamiento de información para realizar inferencias sobre el tiempo. Se propone una búsqueda secuencial en la memoria, siendo generada cada búsqueda por una formulación particular de la respuesta buscada. De hecho, aunque no aparece aquí, Clark hace uso de un formalismo para expresar la estructura profunda de las sentencias que se acerca a una representación interna de un sistema de procesamiento de información.

Al igual que en el ejemplo sobre memoria inmediata, eva-

luar la significación exacta del modelo de Clark nos apartaría de nuestros propósitos. Funciona muy bien en la explicación de los datos, pero sería necesario discutir teorías alternativas, tales como las que usan una imagen interna uni-dimensional sobre la que se representa cada sentencia (HUTTENLOCHER, 1968). Lo que aquí nos importa es que el modelo es un ejemplo de la orientación de procesamiento de información, aplicada a otro aspecto de la psicología.

Aprendizaje verbal y percepción

El espacio y la preocupación por la paciencia del lector aconsejan no continuar a través de otras áreas donde la orientación de procesamiento de información ha producido un trabajo importante. En el campo del aprendizaje verbal (que por razones históricas se desarrolla al margen de la psicolingüística) existen algunos modelos, tales como el EPAM (SIMON y FEIGENBAUM, 1964), que se plantean como programas muy similares a los modelos de solución de problemas. Sin embargo, el cambio hacia el procesamiento de información puede percibirse con más claridad en el trabajo experimental que se está realizando actualmente: por ejemplo, en el uso de mediadores internos, tales como imágenes y procesos memorísticos, en el recuerdo de listas de material verbal (PAIVIO, 1971). Es bien sabido que recordamos mejor si utilizamos una lista sobre-aprendida —“uno es un umbral, dos es un doncel, tres es un tren, ...”—, estableciendo para cada elemento que debe aprenderse una asociación visual clara y distintiva con el objeto nombrado. El paradigma del aprendizaje S — R, donde simplemente se construyen asociaciones de distinta fuerza entre estímulo y respuesta, no permite fácilmente el estudio de tales fenómenos —y, además, tales estudios no han sido realizados. Dada una perspectiva de procesamiento con su interés por la representación interna y su manipulación, tales investigaciones parecen ahora naturales.

El estudio de la percepción es mucho más especializado,

como ya señalé anteriormente, y surgió por el interés en la existencia de un “almacén de imágenes”, principalmente tal como lo planteó SPERLING (1960). Se han convertido ya en relevantes otros muchos tópicos que tienen su lugar en la percepción, como el enmascaramiento visual, de forma que es considerable la cantidad de trabajo realizado. Una colección reciente de Ralph HABER, *Information Processing Approaches to Visual Perception* (1969), pone de manifiesto muy claramente este campo.

Inteligencia artificial como psicología teórica

Hasta aquí hemos prestado atención preferente al estudio del hombre como procesador de información. Ahora es necesario poner de manifiesto las conexiones entre este estudio y la tercera concepción de inteligencia artificial, que habíamos titulado psicología teórica.

A estas alturas debe ya haber quedado claro que la noción organizativa importante es la de un sistema de procesamiento de información, no la de un sistema de inteligencia artificial. Para nuestro propósito, todo sistema del último tipo lo es del primero, pero no a la inversa. (De hecho, no todos los sistemas de inteligencia artificial son sistemas de procesamiento de información, en el sentido en que nosotros utilizamos el término —sistemas simbólicos discretos—, pero hemos puesto gran cuidado en limitar nuestra consideración a estos últimos tipos de sistemas). La inteligencia artificial explora ciertas tareas, como vimos en nuestra discusión del primer punto de vista. Pero muchos aspectos de la conducta humana, que están actualmente bajo estudio, no incluyen tales tareas. El trabajo sobre memoria inmediata se desarrolla dentro de los límites de un sub-sistema particular, y una gran parte del trabajo en psicolingüística consiste en intentar evaluar la naturaleza de las realizaciones gramaticales. Las tareas utilizadas en el estudio del aprendizaje verbal (aprendizaje y recuerdo de listas de palabras y sílabas sin sentido) no tienen un interés directo como tareas de inteligencia artificial. Los programas construidos para

semejantes tareas, tales como el EPAM, están diseñados como teorías psicológicas, no para demostrar que al menos un mecanismo puede realizar esas tareas. De esta forma, la totalidad del estudio de los sistemas de procesamiento de información constituye una psicología teórica —si se concibe al hombre como un sistema de esa índole.

Al comienzo de la presentación de esta tercera concepción, manifesté que los lenguajes teóricos, o clases de sistemas, son una parte del campo científico total. Los sistemas de procesamiento de información constituyen uno de tales lenguajes; pero existen muchos otros. El modo de buscar una teoría sobre un fenómeno determinado, consiste en investigar dentro de una de estas clases —esto es, escribiendo en uno de estos lenguajes. Los comentarios del párrafo anterior implican que la clase relevante para la psicología es el procesamiento de información y no el subdominio denominado inteligencia artificial.

Cada una de las clases de tales sistemas, si están suficientemente bien definidas, origina un estudio independiente. Existe una matemática de ecuaciones diferenciales en la que puede verse con claridad la marca de su estrecha afinidad inicial con la física clásica. Pero también es un campo autónomo. Lo mismo puede decirse de los procesos de Markov y de la clase de procesos estocásticos que subyacen a las llamadas teorías estocásticas del aprendizaje. Cuando la clase permanece claramente mal fornada, como en la metáfora dramática que subyace a la teoría freudiana, no encontramos un desarrollo independiente. Pero hasta la concepción del conocimiento humano como un conjunto de fuerzas recibió un marco adecuado en la teoría de grafos (HARRY, NORMAN Y CARTWRIGHT, 1965).

La inteligencia artificial, en esta concepción, es precisamente una psicología teórica porque se ocupa del estudio de los sistemas de procesamiento de información que solucionan problemas. Tiene parcelas que son bastante fiel reflejo de la psicología (por ejemplo, GPS y EPAM), y otras partes que son decididamente autónomas (por ejemplo, REF — ARF). La contribución de las últimas, en esta tercera concepción, consiste en que desarrollan la

ciencia de los sistemas de procesamiento de información, de forma que se pueden construir y analizar mejor otros sistemas específicos más relevantes para los problemas psicológicos.

Sumario

Hemos presentado tres concepciones distintas de la inteligencia artificial. Es evidente que no tienen un carácter exclusivo. Las actuales investigaciones que gran parte de la comunidad científica denomina inteligencia artificial, podrían desarrollarse de tal forma que prosperaran las tres concepciones o que una llegase a ser dominante.

Existen también otros puntos de vista, y mi selección de los tres analizados no significa que ignore esos otros, aunque indica las orientaciones que me parecen más relevantes en la interpretación del campo. Por ejemplo, uno puede pensar que este campo es propio de la ingeniería, oponiéndolo de alguna manera a la ciencia. Nos desviaríamos demasiado si nos ocupáramos ahora de esta distinción. Además, remito al lector a un libro de mi colega Herbert SIMON sobre las *Ciencias de lo Artificial* (1969), para una perspectiva que justifique el que se considere a campos tales como los computadores, el diseño, la optimización, la inteligencia artificial, la economía y la psicología, como ciencias que guardan un parecido familiar que las pone aparte de las ciencias naturales tales como la física, la química y la astronomía.

Hemos dedicado mucho espacio a la descripción de estas tres orientaciones o concepciones, tratando de ilustrar mediante ejemplos y segmentos de programas parte de su contenido y alcance. Debemos ahora regresar a los problemas de la mente, para valorar la contribución que la inteligencia artificial puede aportar a su solución.

CUESTIONES SOBRE LA NATURALEZA DE LA MENTE

Los problemas surgen desde un punto de vista —a partir de algo que ayuda a estructurar lo que es problemático, lo que merece la pena preguntar y lo que constituye una respuesta (o progreso). No es que el punto de vista determine la realidad, sino sólo lo que aceptamos de la realidad y cómo la estructuramos. Soy lo suficientemente realista como para creer que, a la larga, la realidad pone en juego sus propias oportunidades para aceptar o rechazar nuestros diversos puntos de vista. Podríamos adoptar los puntos de vista recién expuestos y preguntarnos qué cuestiones suscitan acerca de la naturaleza de la mente que valga la pena plantear. Pero, como manteníamos al principio, en este ensayo la inteligencia artificial debe considerarse como un medio y los problemas deben buscarse en otro lugar. ¿Dónde?

Pero surge una dificultad, que descansa en lo que KUHN (1962) planteó y que hemos ido manifestando implícitamente a lo largo de este escrito. “En otro lugar” quiere decir desde otra concepción u orientación —posiblemente desde la filosofía—, como también podríamos decir “en otros lugares”, puesto que las perspectivas acerca del hombre son múltiples. Cada perspectiva plantea sus propios problemas. Las distintas perspectivas se comunican poco entre sí. Ocasionalmente, sin duda, hablan acerca del mismo asunto y entonces es posible la comparación, pero no muy a menudo ni según quisiéramos.

De todas formas, lo intentaré. Voy a considerar un reducido número de tales cuestiones. Cabrá aludir al menos a unos cuantos puntos de contacto y añadir unas pocas clarificaciones que es necesario realizar antes de que nos veamos forzados a re-

gresar a la posición básica: la inteligencia artificial habla sobre la naturaleza de la mente principalmente desde el interior de su propia perspectiva.

PROBLEMAS SOBRE LA NATURALEZA GENERAL DE LA MENTE

Existe una tradición filosófica, en la actualidad totalmente profesionalizada, que plantea cuestiones sobre la mente y sobre nuestro conocimiento de la mente. Consideremos una pequeña muestra:

Llego a la conclusión de que el conflicto central del problema mente-cuerpo es la naturaleza lógica de las leyes de correlación que conectan las cualidades sensoriales directas con los procesos neurofisiológicos (FEIGL, 1963).

Dos cuestiones surgen de inmediato sobre la "mente" tal como se emplea en estas frases. Una es a qué tipo de entidad sustantiva se piensa designar en ellas con el término "mente". La otra es si existe alguna entidad sustantiva del tipo concebido (DUCASSE, 1965).

La cuestión que plantea el problema mente-cuerpo puede establecerse ahora de la siguiente manera: la interpretación que realizamos, o que no realizamos, de la conducta y la experiencia humana, ¿necesita conceptos que no puedan ser analizados en términos de los conceptos utilizados en la interpretación de acontecimientos físicos? (BELOFF, 1965).

Pero aun cuando hubiese estado pensando en voz alta con una cinta magnetofónica cerca, el monólogo resultante hubiese sido difícil de interpretar. ¿Por cuánto tiempo habría de tomar en cuenta una posibilidad para que ésta contase como una decisión? ¿Y con qué grado de firmeza? La primera dificultad para cuantificar lo mental, por tanto, es la de identificar unidades con vistas a su enumeración (QUINTON, 1965).

¿Qué es lo que puede decir la inteligencia artificial sobre todas estas cuestiones? Depende, naturalmente, de cuál sea la concepción de inteligencia artificial que vayamos a adoptar. Desde el punto de vista que la considera ciencia de los métodos-débiles, la inteligencia artificial no dice nada al respecto, puesto que se centra por completo en la relación de los procedimientos con las tareas. Desde la perspectiva de orientación de la exploración de tareas intelectuales, puede decirse alguna cosa, si es que la investiga-

ción filosófica se ha desarrollado lo bastante para concentrar su atención en la realización de algunas funciones intelectuales particulares como criterio para decidir la existencia y funcionamiento de la mente. Pero éste, según veremos más adelante, no ha sido el caso.

Desde el punto de vista de la inteligencia artificial considerada como psicología teórica aparecen respuestas claras:

La mente es un sistema de procesamiento de información.

Las mentes (sistemas de procesamiento de información) están realizadas en sistemas físicos.

La brevedad y simplicidad de estas respuestas revela de nuevo el truco del prestidigitador —todo se esfumó antes de empezar. Es decir, una vez que se adopta la perspectiva del procesamiento de información, estas respuestas se siguen (más o menos) automáticamente. Ello supuesto, las cuestiones que se acaban de consignar o bien son preguntas directas que tienen respuestas directas, o bien son preguntas científicas cuyas respuestas varían según el estado del problema. Por ejemplo, la teoría de la solución humana de problemas en criptoaritmética anteriormente expuesta responde directamente a las preguntas de Quinton —y casi les ofrece solución.

Pero esto no es totalmente satisfactorio, puesto que no hay ninguna necesidad de aceptar este punto de vista. Se trata simplemente de una orientación científica particular, cuyos datos son los esquematizados en la primera parte de este trabajo. No hay al respecto ninguna necesidad forzosa a nivel filosófico. Quizá se pueda plantear más claramente la situación observando que el procesamiento de información es una forma particular de la hipótesis de identidad propuesta por FEIGL (1963). Pero nosotros no hemos llegado a ella a través de ninguno de los tipos de argumentación presentados en su dilatada monografía o la resultante avalancha de trabajos (por ejemplo, MAXWELL, 1966). Ni nos hemos ocupado de ninguno de los problemas que surgen en su seno. ¿Qué es lo que, en función de estos problemas, diferencia a una teoría de procesamiento de información sobre la actividad mental de cualquier otra teoría psicológica? Debe admitirse que,

incluso en la cita de Quinton, sus planteamientos no son sino un eslabón en la búsqueda de problemas filosóficos más profundos. La respuesta a estas cuestiones sólo desplazaría el punto de partida de la investigación.

Argumentos analógicos

Es posible que se pueda obtener ayuda de otra forma: es decir, produciendo analogías que permitan ver las cuestiones planteadas a una nueva luz. El mejor ejemplo que conozco aparece en un ensayo de PUTNAM (1960), donde éste muestra que todas las cuestiones consideradas en la filosofía de la mente se pueden plantear en relación con las máquinas de Turing, teniendo en cuenta que dichas máquinas poseen tanto efectores como receptores (para observación externa) y un estado interno (estado entendido en sentido técnico). Pretende demostrar que en uno y otro contexto estas cuestiones tienen las mismas respuestas y por las mismas razones. El argumento es analógico, pero fructífero.

Se puede ofrecer un argumento analógico similar en el sentido de que:

La neurofisiología es casi totalmente irrelevante para la naturaleza de la mente.

Tal argumento se apoyaría en la proposición análoga de que:

La electrónica es casi totalmente irrelevante para la naturaleza de la inteligencia artificial.

La última proposición es un lugar común en las ciencias del computador. Los computadores podrían estar igualmente contruidos de acuerdo con uno cualquiera de entre muchos tipos de tecnología. La conducta de la máquina viene especificada por el manual de programación, y esto está totalmente realizado fuera del “hardware” subyacente.

Esto es cierto, como todos sabemos, porque pueden existir mecanismos funcionalmente equivalentes. Pero este no es el argumento completo. Porque si las exigencias de la tecnología

subyacente son demasiado complejas y complicadas, entonces puede ser muy difícil inventar o descubrir mecanismos funcionalmente equivalentes. Pero en la construcción de computadores sólo se necesitan unas pocas propiedades generales para obtener una máquina capaz de procesar información arbitraria: una memoria simbólica amplia y fiable; operaciones de lectura, impresión y direccionamiento efectuables en ella; y un conjunto de modesto tamaño de estados internos que el organismo puede asumir como una función de los símbolos que lee y que conducen a la realización de las operaciones adecuadas. Esto es, desde luego, algo que debemos pedir a un mecanismo. Pero de hecho no es gran cosa, y hay una gran cantidad de formas de conseguirlo. Los medios por los que se consigue orientan algunos aspectos del sistema: rapidez, memoria, tamaño, fiabilidad, etc. Esto explica la frase "casi totalmente irrelevante", en lugar de "totalmente irrelevante". Pero estas características son globales. El detalle de la tecnología no indica gran cosa (véase también NEWELL y SIMON, 1972, Cap. 14).

Dicho de otro modo, y continuando con el argumento, si la inteligencia artificial (considerada ahora como exploración de tareas intelectuales) produce realizaciones similares a las de la mente (sin que nos preocupe la exactitud del parecido con la mente ni en qué forma), y esta conducta sólo depende de unas cuantas características generales del sistema físico subyacente, ¿no podría ocurrir que la conducta mental de los organismos biológicos sólo dependiera de unas cuantas características generales del mecanismo biológico subyacente? El argumento es analógico. Y pienso que fructífero.

Obsérvese que estos dos argumentos, tanto el de Putnam como el de la neurofisiología, no son argumentos extraídos de la inteligencia artificial, sino que surgen de la naturaleza general de los sistemas de procesamiento de información. El único uso de la inteligencia artificial consiste en ayudar a garantizar la relevancia de la analogía en el segundo argumento; no se utilizó en absoluto en el de Putnam. Esto guarda un paralelo con el constructo subyacente de la concepción de la inteligencia artificial como psicología teórica, que consiste en sistemas de procesamiento de información y no en sistemas de inteligencia artificial *per se*.

CUESTIONES SOBRE LA MENTE DE LOS COMPUTADORES

El computador mismo se ha convertido en un objeto de investigación filosófica. Consideremos algunos ejemplos:

Propongo considerar la cuestión "¿Puede pensar una máquina?" (TURING, 1950).

¿Cómo podríamos reconocer una *entidad* (por utilizar un término que no prejuzga el problema de si es un organismo o una máquina) como entidad que tenga una mente? (ANDERSON, 1964).

¿Existe una diferencia esencial entre un hombre y una máquina? (SCRIVEN, 1953).

¿Podría un robot tener sentimientos? (ZIFF, 1959).

¿Tendrían los robots derechos civiles? (PUTNAM, 1960).

Más seriamente, podemos preguntarnos si existe algún fundamento en la revelación cristiana para pontificar allí donde fracasa la deducción. ¿No es teológicamente imposible la producción de un ser humano artificial, puesto que usurparía la prerrogativa divina? (MACKAY, 1965).

Con seguridad, el trabajo sobre inteligencia artificial es aquí relevante. Por ejemplo, el artículo de TURING (1950) y la prueba que propone son muy discutidos tanto en la literatura filosófica como en la de inteligencia artificial. Pero la discusión discurre en direcciones completamente distintas en cada caso. En inteligencia artificial conduce a la concepción de que ésta es exploración, y a los intentos de programar los computadores para que realicen funciones intelectuales cada vez más arduas. En filosofía, la discusión lleva a saltar por encima de todo cuanto es relevante para la inteligencia artificial y a preguntarse qué se podría conocer con certidumbre, aun cuando se confirmase el planteamiento empírico. Este salto debe ser ejemplificado con detalle, porque me parece que constituye el núcleo de la irrelevancia de la inteligencia artificial para los problemas filosóficos de la mente.

Cuando se dice que es imposible que una máquina sea consciente, no siempre está claro en qué medida se entiende que es una objeción lógica y en qué medida es empírica. Se puede pensar que es insalvable algún obstáculo práctico determinado, como por ejemplo proporcionar a una máquina un medio de comunicación. Pero debe recordarse que es *concebible* superar tales obstáculos empíricos: y un argumento fundamentado únicamente sobre estas diferencias no es lógicamente demostrativo. Esta discusión no considera que sean diferencias esenciales *ninguna* de tales diferencias prácticas: así, las máquinas que puedan compararse con el hombre deben ser capaces de hablar, gesticular, pasear, etc., al igual que ser sensibles a la luz, al sonido y a otras condiciones ambientales. Se hará referencia a ellas como robots. Un robot puede hacer cualquier cosa que pueda hacer una máquina. Surge entonces la segunda cuestión: ¿existe alguna diferencia esencial entre los robots y los seres humanos? (SCRIVEN, 1953).

Deseo plantear el tipo correcto de robots. Deben ser autómatas y sin duda máquinas.

Supondré que esencialmente son máquinas computadoras, que tienen microelementos y que pueden necesitar cualquier tipo de micromecanismos para el funcionamiento de estos prodigios de la ingeniería. Además, voy a suponer que se alimentan de baterías microsolares: en lugar de tomar comida, tomarán luz.

Y si es evidente que nuestros robots son sin duda máquinas, entonces en todos los demás aspectos deben ser tan semejantes al hombre como se desee. Pueden ser del tamaño de los hombres. Cuando estén vestidos y enmascarados, pueden ser virtualmente indistinguibles de los hombres en prácticamente todos los aspectos: en su apariencia, en su movimiento, en las expresiones que emitan, etcétera. Así, salvo por las máscaras, cualquier hombre corriente podría tomarlos por hombres ordinarios. No sospechando que fuesen robots, nada en ellos le haría sospechar.

Pero desenmascarados, los robots se mostrarían con todo su brillo metálico. Lo que aquí se pone en cuestión no es si podemos desdibujar la línea entre un hombre y una máquina, y atribuir así sentimientos a la máquina. La cuestión es si podemos atribuir sentimientos a la máquina y desdibujar así la línea entre un hombre y una máquina (ZIFF, 1959).

¿Qué se puede ganar de los *hechos* de la inteligencia artificial cuando ésta se da por supuesta? Quizá un incentivo para considerar tales cuestiones, pero no una ayuda para el argumento.

El término simulación aparece repetidamente en la literatura sobre el computador y la mente, aunque sucede que no aparece en las citas anteriores. Tal vez haya que decir algo aquí, aun-

que sólo sea una clarificación de uso y no nada sustantivo —que es lo que le otorgaría importancia. Si se adopta el punto de vista de que la inteligencia artificial es psicología teórica, la simulación (la ejecución de un programa que se propone representar alguna conducta humana) es simplemente el cálculo de las consecuencias de una teoría psicológica. La relación que guarda con su objeto empírico es exactamente igual a la de cualquier otra teoría. De hecho, en el mundo de la computación, el término simulación se utiliza indiferenciadamente (y con el mismo significado) para hacer referencia a las simulaciones de tareas empresariales, soluciones de problemas, reactores nucleares, estructuras de tráfico y ecología. No hay necesidad de desarrollar con detalle el argumento. En un reciente estudio, FODOR (1968) dedica por entero un capítulo final a la naturaleza de la simulación y llega exactamente a la misma conclusión y por la misma razón (véase también SIMON, 1969):

Hasta aquí nos hemos interesado por la siguiente cuestión: “¿Cuándo, si es que alguna vez, contaría la simulación de la conducta orgánica mediante una máquina como una explicación de esa conducta?”. He sugerido que esta cuestión podría entenderse adecuadamente como un caso especial de esta otra: “¿Bajo qué circunstancias la capacidad de una teoría para explicar los datos relevantes de observación, la convierte en una teoría verdadera?” (FODOR, 1968).

Existen otros problemas de clarificación, como por ejemplo la significación del teorema de Gödel, o la significación del enunciado “los computadores sólo hacen lo que se les dice que hagan”. (En ninguno de ambos casos se trata de una significación profunda). Pero estas cuestiones han sido tratadas frecuentemente y bastante bien. Me parece que es innecesario reiterar los análisis bien establecidos. Habremos de orientarnos hacia otras direcciones si deseamos obtener algo de la inteligencia artificial en relación con el estudio de la mente.

CUESTIONES SOBRE LA NATURALEZA DE LAS OPERACIONES MENTALES

En el desarrollo de las investigaciones sobre inteligencia artificial, la atención se dirige hacia diversos tipos de operaciones específicas —hacia la búsqueda, la inducción, el descubrimiento. En la literatura filosófica se presta igual atención a las operaciones mentales. A menudo, especialmente en las décadas recientes, estos tratamientos están saturados de un análisis del uso del lenguaje, como si fuera posible utilizar el sentido lingüístico plasmado en una comunidad lingüística como fuente de datos objetivos. Así, Anscombe escribe,

¿Por qué no puede un niño de seis meses fingir que siente dolor?
(ANSCOMBE, 1958).

Creo que nuestra autora no quedaría satisfecha con una indagación en la línea de nuestro ejemplo de criptoaritmética (dejando de lado el hecho de que, de momento, no está cercano un análisis específico de tal tipo). Casi la mitad de la tarea tendría que basarse en un análisis de lo que entendemos por “fingir”. El modelo del niño no haría referencia de por sí a esa cuestión, sino sólo a la manifestación de la conducta en términos de sus generadores simbólicos. Podría llevar incorporado un concepto análogo a fingir, si es que, ciertamente, un niño de seis meses pudiese fingir. Pero ésta sería una definición en sus propios términos y no un reflejo de las leyes de uso en la comunidad lingüística. Tales definiciones podrían producir un impacto en los intereses del filósofo, puesto que plantean alternativas que pueden reemplazar a las lingüísticas (al igual que la física adoptó el término “energía”, cuyo significado primario controla ahora). Pero para términos de la sutileza que atrae a los filósofos, tal posibilidad hay que dejársela al futuro. En los momentos actuales no se puede decir nada sobre “fingir”, ni en los niños ni en ninguna otra cosa.

Inducción. Debemos buscar en consecuencia algunas operaciones que la inteligencia artificial haya estudiado y que la filosofía haya intentado comprender directamente. La inducción es un candidato evidente. Pero, una vez más, al examinar más de cerca la cuestión, nos encontramos con que las dos se han interesado por la inducción en forma bastante distinta. La inteligencia artificial se interesa en averiguar cómo se producen realmente las inducciones —y en la comprensión de inducciones todo lo complejas que sea posible. La filosofía se interesa por la justificación de la inducción. De esta forma:

En la discusión filosófica de la inducción, un problema ocupó por largo tiempo el centro del escenario —hasta el extremo, por cierto, de que se lo conoce como el problema de la inducción. Es el problema de justificar la forma en que fundamentamos nuestras creencias y afirmaciones sobre asuntos empíricos, tanto en la investigación científica como en nuestras ocupaciones diarias, mediante evidencia lógicamente incompleta (HEMPEL, 1966).

Seguramente no es de interés examinar la manera en que los métodos como el de Hipótesis y Comparación realizan inducciones de cierta dificultad. Es evidente que no proporcionan ninguna sugerencia nueva sobre el problema de la justificación. Una vez proporcionado el espacio, el juego está ya casi terminado, y no hay justificación (en el sentido de certidumbre) para los espacios utilizados. Además, puesto que los espacios han sido construidos en el momento del diseño, uno puede inclinarse a decir, desde un punto de vista filosófico-crítico, que los programas que se ajustan a este método no son, en absoluto, programas inductivos. Sin embargo, este es un razonamiento peligroso, porque ¿qué sucede si construimos a continuación un programa cuyo espacio de hipótesis esté compuesto por todas las gramáticas BNF (FELDMAN, 1972)? ¿Es que al proporcionar al programa la información de que la hipótesis es expresable en *alguna* BNF se está realizando todo el trabajo inductivo?

Un candidato aún más evidente que la inducción parecerían serlo la demostración y la lógica deductiva. Pero el esfuerzo masivo por comprenderlas es anterior a las ciencias del computador, y por tanto a la inteligencia artificial. (Este esfuerzo consti-

tuye en realidad una gran parte del capital intelectual empleado para el desarrollo de las ciencias del computador). En cualquier caso, la deducción y la prueba demostrativa son actualmente consideradas como algo que está más allá del concepto de mente, la cual sólo puede reflejar su bien comprendida estructura.

Se da en filosofía una curiosa asimetría entre la forma en que se está tratando a la inducción como problema y la forma en que fueron tratadas la deducción y la prueba demostrativa. En el caso de la deducción no sólo existía el interés por la formalización, por el rigor y por los fundamentos de la inferencia. También existía interés por saber si los modelos de lógica deductiva que se estaban creando eran suficientes para las tareas deductivas del mundo —particularmente para las de la matemática y la ciencia. Siempre ha sido importante para la lógica deductiva demostrar su adecuación. Esto impulsó el desarrollo de sistemas lógicos no triviales, dentro de los cuales la obra de WHITEHEAD y RUSSELL (1910) constituye un monumento clásico.

Sin embargo, con una sola excepción, parece que esto no ha ocurrido nunca en el caso de la inducción. En este campo se opera con una serie de ejemplos simples (aunque ingeniosos) de inducciones que constituyen el núcleo de los diversos enunciados de la justificación. Pero no hay la menor preocupación por averiguar si los distintos tipos de inducción que inquietan a la filosofía son suficientes para las tareas inductivas del mundo. Ni siquiera se sabe si esos tipos de inducción abarcan el espacio de lo que se necesita. ¿Qué tipos de inducción se utilizaron para derivar la tabla de Mendeleiev? ¿O para derivar de la serie de Balmer el modelo de Bohr? ¿O para derivar el término siguiente en un test de inteligencia de terminación de series?

La única excepción de que tengo noticia es la probabilidad, que es una forma de inferencia inductiva. La fundamentación de la probabilidad ha sido investigada de una manera similar a como lo ha sido la fundamentación de la prueba deductiva. Esto sugiere que la dificultad radica en que la filosofía misma no está preparada para proporcionar la explicación del campo de la inferencia inductiva. Si se introduce desde fuera el desarrollo de una

técnica de inferencia, como ocurrió en matemática, en la ciencia formalizada, la probabilidad y la estadística, entonces lo que la filosofía añadirá a esta materia es el problema de si sus conceptos fundacionales son suficientes para la tarea en cuestión. Sin una materia de este tipo introducida por vía independiente, simplemente no surgirá el problema.

Las especulaciones del párrafo anterior, realizadas por alguien que está fuera de ese campo, sólo tienen la intención de hacer plausible la siguiente sugerencia. Una posible contribución de la inteligencia artificial consiste en ocuparse de los detalles de muchos casos de inducción compleja, hasta poder aclarar mejor en qué consiste su estructura de información. Ello podría permitir un contacto con la filosofía de la inducción —aunque esto, naturalmente, no pase de ser una especulación adicional. Está fuera de los límites de esta sección analizar con detalle tales implicaciones en lo que respecta al método de Hipótesis y Comparación. Pues si bien es claro que las operaciones mentales son un aspecto de la mente, también lo es que este tipo de preocupaciones nos está apartando de los conceptos centrales de la mente.

CUESTIONES SOBRE LA NATURALEZA DE LOS SISTEMAS DE PROCESAMIENTO DE INFORMACION

Una cosa ha quedado, tal vez, establecida. No es fácil aplicar el contenido de una orientación para responder a los problemas de otra. Tomando a propósito el caso de la filosofía como materia de ensayo, lo hemos intentado de diversas formas. Mi propia conclusión —no puedo hablar por el lector— es que o bien las muestras escogidas fueron insuficientes, o bien las respuestas ahogaron los problemas reales (si se lo mira desde la otra orientación).

¿Qué problemas deberíamos plantear sobre la naturaleza de la mente? Parece evidente, por ahora, que deben ser los sugeridos desde el seno de la orientación de procesamiento de información. Simplemente, identificamos el concepto de mente con el de un sistema de procesamiento de información. Uno pudiera sentirse tentado a insistir en que de alguna manera se trate de un sistema muy especial de procesamiento de información —que no sería deseable utilizar el término mente en relación con la abigarrada diversidad de sistemas de procesamiento de información (por ejemplo, para un sistema de contabilidad de nóminas). Pero esto nos retrotrae de nuevo a la cuestión de saber si las nociones determinantes son las formas en que hemos utilizado el término desde el punto de vista de otras concepciones.

Tales problemas no se plantean dentro de esta concepción. Las cuestiones adecuadas que interesan aquí se refieren a mecanismos y organizaciones de procesamiento que permitan ampliar las clases de realizaciones. Las otras cuestiones no están todavía lo bastante maduras como para que puedan ser respondidas, y uno de los principios filosóficos que gobiernan el trabajo cotidiano del científico consiste en no preguntar una cuestión antes de tiempo.

Deseamos plantear cuestiones sobre sistemas de procesamiento de información que estén listas para ser planteadas. Consideremos, a este respecto, dos candidatos: la naturaleza de la generalidad y la naturaleza de un problema mal estructurado. No son problemas nuevos. Yo he trabajado ya algo en ellos (ERNST y NEWELL, 1969; NEWELL, 1969) y lo que voy a decir se inspira bastante en esos esfuerzos previos. Pero estoy seguro de que son problemas abiertos, tanto para mí como para el campo de estudio. Presiento que dentro de poco conseguiremos desarrollar respuestas adecuadas, aunque no necesariamente relacionadas con los aspectos que aquí discutiremos.

Los problemas se derivan, en principio, de la concepción de la inteligencia artificial como exploración. Porque inciden sobre dos aspectos de la solución de problemas que para nuestra aún insegura mirada parecen difíciles de manipular mediante mecanismos. El campo de estudio de la inteligencia artificial reconoce en toda su extensión la importancia de estas dos cuestiones, aunque ciertamente aquí no haré más que someter el problema a discusión.

¿Cuál es la naturaleza de la generalidad?

A nivel informal, se distingue fácilmente la potencia de un programa por su generalidad. Los programas X e Y son ambos programas de juego de ajedrez, pero Y vence siempre a X; de esta forma, Y es mejor que X, pero no más general. Además, se asegura a menudo que podemos construir programas para hacer cualquier tarea aislada tan bien (o mejor) que un humano, pero que no podemos construir un programa que haga todas las tareas. Tal pensamiento expresa la creencia de que los programas deben ser considerados *ad hoc* para cada tarea.

En inteligencia artificial es fácil distinguir los esfuerzos por la generalización de los esfuerzos por la potencia dentro de una tarea determinada. Los problemas de ajedrez y de damas no tienen pretensiones de generalidad. Programas tales como GPS y

REF — ARF, se centran claramente sobre la generalidad, como lo indican las pruebas de teoremas del cálculo de predicados.

También es necesario separar la generalidad de la cantidad de especificación —es decir, de qué información se da sobre el problema. Todos sabemos qué valor tiene cualquier indicación o sugerencia que se nos dé además de la definición “usual” de un problema. Un ejemplo extremo al respecto lo constituyen las formas usuales de las máquinas universales de Turing, donde una máquina puede realizar cualquier cosa que pueda hacer otra máquina, pero sólo si se le suministra una descripción total de lo que haría la otra máquina. Los únicos poderes específicos que debe tener la máquina universal, son algunos efectores terminales equivalentes (en el caso de las máquinas de Turing, escribir ceros y unos en una cinta) y la capacidad de leer la descripción, de forma que pueda simular la conducta.

No existen procedimientos moderadamente satisfactorios para definir la generalidad (o la dificultad, o la especificidad). En la práctica, han sucedido dos cosas. En primer lugar, está el juego de “cualquier cosa que pueda hacer tu programa, lo puede hacer el mío —y el mío hace además otras cosas”. De esta forma, cada programa de propósito general tiene que abarcar el mismo campo que sus predecesores. En segundo lugar, la generalidad está vinculada a la existencia de un lenguaje, mediante el cual se puedan describir los problemas al sistema. Cualquier programa que sólo tenga un conjunto determinado de parámetros (como en una gran subrutina) simplemente no es candidato para conseguir generalidad. Por ejemplo, REF — ARF sí lo es, en parte porque tiene un lenguaje de entrada bien definido. Y de hecho no es posible enumerar de forma significativa los problemas que pueden plantearse a ARF (el solucionador de problemas) a través de REF.

Aunque indudablemente es una condición necesaria, el poseer un lenguaje no nos lleva muy lejos. En el momento en que los lenguajes de entrada tienen la potencia expresiva del cálculo de predicados de primer orden, entonces el alcance total del lenguaje no proporciona ya indicios de cuán general sea un programa. Muchos de los programas para probar teoremas del cálculo de

predicados de primer orden son ya perfectamente generales, pero esto dice muy poco. Decir que un programa probador de teoremas puede hacer integración simbólica, como lo hacen el SAINT de Slagle y el SIN de Moses, no sirve para especificar las distinciones relevantes. Esto demuestra bastante bien que los criterios de calidad no han sido realmente purgados de la noción de generalidad. Sin embargo, tal como observamos anteriormente al discutir el programa de Waldinger y de Lee (y también el de Green), se está extendiendo bastante el alcance real de los probadores de teoremas del cálculo de predicados.

No valdría la pena considerar con demasiada detención las dificultades de obtener definiciones y medidas de generalidad. Los problemas realmente interesantes de la generalidad radican en intentar determinar los mecanismos que puedan alcanzarla. Existen diversas hipótesis sobre la naturaleza de un eventual solucionador general de problemas.

Un único método general de búsqueda

Los mejores ejemplos son probablemente el GPS y el MULTIPLE (SLAGLE, 1965), aunque en algunos aspectos también es adecuado REF — ARF, como ocurre con todos los probadores de teoremas del cálculo de predicados. Se adopta una única formulación muy general de un problema. El GPS concibe todas sus tareas como objetos que pueden ser transformados en nuevos objetos por medio de un conjunto de operadores. Los probadores de teoremas conciben la tarea como un conjunto de cláusulas, con una regla única de inferencia (Resolución, Hiper-resolución, etc.) que genera nuevas cláusulas para el conjunto. Invariablemente, los métodos utilizados son búsquedas de algún tipo, por lo general una forma de búsqueda heurística (por lo demás, es difícil ver qué otra cosa podrían ser).

El Gran Conmutador (Experto)

Los expertos solucionan problemas, según se dice, porque saben lo que hay que hacer; la investigación sólo se debe buscar como último recurso. Un solucionador general de problemas es un programa que es (sea lo que sea) experto en muchos campos —lo que significa que tiene conocimiento especializado en la solución de problemas de múltiples campos. Así, la estructura de un solucionador general de problemas comprende un programa de diagnóstico (un “gran conmutador”) que observa un problema determinado y decide qué programa (método) especial debe aplicarse. Los métodos, por su parte, están especializados y no pueden transferirse de una tarea a otra.

Un ejemplo de programa deliberadamente diseñado con esta filosofía es el SIN, un programa de integración de MOSES (1968). SIN comprende tres estadios por los cuales ha de pasar un problema, y a su vez cada uno de esos estadios dispone de múltiples métodos: el primer estadio dispone de tres; el segundo de once; y el tercero de dos. Están todos ordenados según el grado de esfuerzo, siendo los primeros los que menos esfuerzo cuestan. Dentro de cada estadio, una rutina de diagnóstico realiza la selección del método adecuado. Cada método se fundamenta en características muy específicas de las expresiones que deben integrarse, y muchos de ellos incluyen la realización de algún tipo de análisis matemático particular.

La integración simbólica es, naturalmente, un campo limitado. Pero es concebible que la estructura de los solucionadores generales de problemas sea la de un SIN muy enriquecido. Un sistema tal necesitaría una memoria muy grande, y todavía sabemos muy poco sobre qué es lo que serían capaces de hacer memorias de grandes dimensiones. Cabe suponer que un sistema con diez mil métodos, por ejemplo, sería un solucionador muy general de problemas, semejante a un mosaico de diez mil piezas encajadas que cubriese una pared (véase CHASE y SIMON, 1972, para hacerse una idea de cómo sería capaz una gran memoria de jugar al ajedrez).

Una hipótesis sobre la generalidad surge de la concepción de la inteligencia artificial como el estudio de los métodos débiles (NEWELL, 1969). Un solucionador general de problemas es aquél cuya colección de métodos contiene una secuencia de ellos que son cada vez más débiles, de forma que no exista ninguna situación que atraviese el solucionador de problemas (en términos de su estado de ignorancia sobre el ambiente de la tarea), donde éste no disponga al menos de un método cuyo enunciado del problema se ajuste a la situación. La diferencia entre esta hipótesis y la del experto es que en este último caso, el campo total aparece repleto de métodos altamente específicos, pero poderosos; en el primer caso, el campo de estudio aparece cubierto por métodos cuyos dominios son cada vez más extensos al tiempo que disminuye paulatinamente su grado de potencia.

Ambas hipótesis requieren algún tipo de actividad de diagnóstico para seleccionar el método. Uno puede ceder a la tentación de localizar en esta selección la clave de la acción inteligente (MINSKY, 1961). Sabemos poco sobre programas de selección de método. Algunos programas existentes de esta índole consisten en redes fijas de discriminación, similares al EPAM (SIMON y FEIGENBAUM, 1964): el GPS realizaba una selección entre 18 métodos distintos mediante un artificio de este tipo. Alternativamente, las estructuras de transferencia condicional están simplemente insertas en el programa ejecutivo. La trivialidad de los ejemplos actuales puede deberse a que carecemos de amplias e interesantes colecciones de métodos para seleccionar. En cualquier caso, parece, al menos, que la selección de programas *debe* ser fundamentalmente simple. Si incluyese un procesamiento muy complejo, entonces es que ya se habría realizado implícitamente una decisión sobre el "método" a utilizar.

Es posible que cada problema, presentado (por ejemplo) mediante una descripción lingüística, origine un esfuerzo de organización que reúna un poder ejecutivo específico para cada problema. Métodos tales como los anteriormente esbozados no pre-existen en el solucionador de problemas como estructuras simbólicas estáticas que haya que interpretar. Por el contrario, son creados *ad hoc* en función de la situación.

Esta es una propuesta atractiva, puesto que parece hacerse cargo del tipo de particularidad discreta que caracteriza, por ejemplo, a la tarea analógico-geométrica discutida en la primera sección. Es difícil creer que una persona esté preparada para la forma particular en la que se le presente un problema: A:B::C:(1, 2, 3, 4, 5,). Lo mismo vale decir cuando se introduce por vez primera a un solucionador de problemas en el ajedrez o en la prueba de teoremas con un conjunto nuevo de axiomas y de reglas de inferencia.

Existen pocos indicios que nos den una pista de cómo pueden tener lugar tales estructuraciones. Un programa reciente de Don WILLIAMS (1969) da un pequeño paso en esta dirección. Trabaja generalmente en problemas de inducción tal como aparecen en los tests de inteligencia. Primero examina los problemas sencillos que aparecen siempre, a manera de ejemplo, al introducir una nueva sección de problemas en tales tests. A partir de esta inspección, construye un programa particular de inducción que manipula la tarea. Es así como produce un ensamblamiento ejecutivo. El Cuadro 1.15 muestra algunos problemas que él puede estructurar por sí mismo y después manipular.

Existen dos limitaciones en la capacidad del programa de Williams para proporcionar claves de cómo se consigue la estructuración ejecutiva. En primer lugar, sólo tiene un método de solución. Reduce todos los problemas a una forma en la que se pueda evocar este método. Tal es exactamente lo que deseamos que haga un ensamblador ejecutivo, pero ello proporciona poca luz sobre la manipulación de una pluralidad de métodos. La segunda limita-

Cuadro 1.15 Tipos de Pruebas ensambladas en el Aptitude Test Taker	
ANALOGIAS DE LETRAS	IJ JI PO OP ED _ _ Rellenar los blancos.
SERIE DE LETRAS	A X B Y A X B Y A X B _ _ _ Rellenar los blancos.
AGRUPACION DE LETRAS	A A B C A C A D A C F H A A C G Señalar la que no corresponde.
GRUPOS DE NUMEROS	35 110 75 Decir qué hay de común.
RELACIONES DE NUMEROS	2 6 3 9 4 12 6 15 Señalar la que no corresponde.
SERIES DE NUMEROS	15 18 21 24 27 30 Descubrir la regla.
CORRECCION DE NUMEROS	1 2 3 4 5 7 Descubrir el erróneo.
DIVISAR TENDENCIAS	ANSIA BANAL CAJA DEDO EXCUSA ¿Cuál es la tendencia?
GRUPOS DE PALABRAS	MAIM TEST GANG LABEL ¿Qué hay de común?
RELACIONES DE PALABRAS	FIESTA SIESTA CAMA RAMA CIELO _ _

ción radica en la variedad de situaciones que puede estructurar. Maneja variaciones en la forma de respuesta, variaciones en el alfabeto y variaciones en los grupos o las secuencias de letras. Esto es suficiente para manipular la mayor parte de los tests. Desgraciadamente, éstos sólo recogen unos pocos dominios de variabilidad. No está claro cómo extender el programa a situaciones más abiertas —lo que es justamente el gran problema. Sin embargo, el programa de Williams nos proporciona un ejemplo no trivial de estructuración ejecutiva.

Estas hipótesis no agotan las posibilidades por las que se puede alcanzar la generalización. No descartan, por ejemplo, que existan métodos aún por descubrir que se relacionen específicamente con la consecución de la generalidad. Tal posibilidad es quizá la principal alternativa a las anteriores hipótesis —que todavía no hemos tocado el problema de la generalidad en nuestras exploraciones.

No deberíamos llegar a la conclusión de que las diversas hipótesis que hemos enumerado hasta aquí son algo exclusivo. La naturaleza de la inteligencia, que implica el uso de significados alternativos, casi garantiza que los sistemas efectivos generales hayan de consistir en mezclas y combinaciones. Trabajos recientes, como por ejemplo el programa construido por WINOGRAD (1972), apoyan esta conclusión. Mantiene una estrecha semejanza con el experto, pues actúa distribuyendo la acción del programa en muchos programas distintos. Cada subprograma debe estar muy adaptado a su sub-tarea, invocando un conocimiento sustantivo especial (o sea, el experto). Sin embargo, el sistema completo se halla instalado encabezando un sistema de programación que proporciona un modo uniforme de evocar metas y de mantener los cuidados y funciones que se asocian con la búsqueda de metas (HEWITT, 1971), incorporando así alguno de los principales componentes de las orientaciones de-un-solo-método de búsqueda heurística.

No existe una separación clara entre los aspectos de diagnóstico del programa y los aspectos de intento de solución; unos y otros se encuentran entrelazados. El sistema de programación

de base dispone también de una capacidad uniforme y poderosa de comparación de estructuras que subyace al sistema de selección de método y permite distribuirlo. (La capacidad de búsqueda de estructuras proporciona también otras facilidades, como por ejemplo un acceso general a la base de datos). El sistema es impulsado por la entrada de sentencias que se introducen desde el exterior. En respuesta a éstas, el programa monta un subprograma en función de la semántica de la sentencia de entrada. Así, aunque de modo un tanto limitado, el programa incorpora también la noción de ensamblamiento de programas.

Se impone decir unas palabras finales sobre el aprendizaje, puesto que con frecuencia se piensa que la generalidad se consigue a través de un programa que aprende. Tal propuesta demuestra simplemente que el problema ha sido mal interpretado (NEWELL, 1962). La organización final debe solucionar el problema, dejando al margen el hecho de que esa organización se consiga mediante el aprendizaje, la instrucción, o la evolución. La insistencia en que la realización de programas debe construirse mediante aprendizaje en ambientes especiales, tales como los que presentan secuencias de ejemplos denominados "éxito" y "fracaso", puede servir para limitar los tipos de programas de realización susceptibles de ser contruidos. Pero lo que necesitamos en estos momentos es generar, como candidatos a tal fin, organizaciones de realización que tengan capacidad de solución general de problemas.

¿Qué es un problema mal estructurado?

La mayor parte de las tareas estudiadas en inteligencia artificial parecen estar bien definidas o bien estructuradas. Algún sesgo se produce en nuestra percepción, naturalmente, y debemos juzgar un programa y su tarea de manera distinta después de unos cuantos años de familiaridad y estudio. Sin embargo, la prueba de teoremas, la realización de juegos, la solución de problemas, y los problemas de las ciencias empresariales matemáticamente formalizadas, se sitúan claramente al final de una escala de buena estruc-

tura —aun cuando la dimensión sea en sí misma vaga.

A modo de ilustración: ¿qué tarea podríamos escoger como mal estructurada? Decidir una profesión. El descubrimiento de una nueva teoría científica. La valoración de un nuevo ballet. Planificar las actividades de un día libre. Pintar un cuadro. La conversación con un viejo conocido. El diseño de una nueva casa. Inventar algo. Hallar un modo de escaparse del Vietnam. Idear un experimento crítico para comprobar una hipótesis científica. Generar la presente lista.

Nada ganaríamos continuando, porque muchas de las actividades mal estructuradas escaparían a nuestra relación.

Se podría esperar, especialmente en la concepción de la inteligencia artificial como exploración, que hubiesen destacado unas cuantas tareas como ejemplos especialmente interesantes de problemas mal estructurados. No parece que haya ocurrido así. Sin embargo, se han lanzado diversas hipótesis sobre el carácter general de los problemas mal estructurados.

1. *Limitaciones abiertas.* REITMAN (1965) fue el primero en intentar profundizar en el enunciado de los problemas mal estructurados. Observó que toda la experiencia con programas heurísticos se concentraba sobre tareas bien estructuradas. Su formulación se centró en la forma en que se definían las tareas. Supóngase que el paradigma de un problema es $(X, \Rightarrow, Y)$, donde X es la información proporcionada, $\Rightarrow$ es la forma de realizar cambios en la situación, e Y es la situación deseada. Entonces, la mala estructuración puede ser localizada en estos diversos aspectos. Los grandes programas de búsqueda heurística, por ejemplo, trabajan sobre problemas donde X e Y están bien definidos: la mala estructuración radica en $\Rightarrow$. Además, si se reconsidera el aserto de que los primeros solucionadores de problemas trabajaban sobre problemas mal estructurados, se hallará que se centraban sobre el modo en que tales problemas no estaban bien definidos.

La idea de Reitman consistió en caracterizar la parte mal estructurada de un problema mediante las condiciones o limita-

ciones que se habían dejado abiertas: es decir, aquellos aspectos del problema que simplemente no estaban especificados en el enunciado del mismo. Estos aspectos quedan “abiertos” y deben ser “cerrados” por el mismo solucionador de problemas antes de que pueda trabajar sobre el problema en cuestión. El enunciado de éste no determina cómo se deben cerrar tales limitaciones. Por el contrario, el solucionador de problemas ha de recurrir a su conocimiento general de muy amplio contexto para poder llegar a una decisión acerca de cómo cerrar estas limitaciones.

Ejemplos simples, pero reales, son los problemas de inducción discutidos anteriormente. El enunciado del problema no establece qué clase de aplicaciones deben utilizarse —está abierto. Para una clase dada de aplicaciones, se plantea el problema bien definido del descubrimiento de aquellas que son consistentes con los ejemplares. Si son posibles muchas aplicaciones, aun dentro de un espacio determinado de ellas, entonces se produce otra limitación abierta, que consiste en cómo seleccionar una en particular. Los solucionadores humanos de problemas recurren a su más amplio conocimiento del mundo para una clase de aplicaciones. Nosotros podemos predecir (por ejemplo, como psicólogos que controlan la práctica de tests) qué clase debe seleccionarse, porque participamos también del mismo contexto amplio y sabemos qué posibilidades están contenidas en él.

Similarmente, la segunda limitación abierta (cómo seleccionar entre múltiples soluciones) también se clausura mediante el recurso a un contexto más amplio. La adopción de algún criterio de simplicidad es casi universal —pero no porque estuviese establecido en el enunciado del problema. El solucionador de problemas debe saber que tal cosa es aceptable en el contexto más amplio donde se plantea el problema. Qué criterio de simplicidad se llega a seleccionar depende del método utilizado para generar posibles aplicaciones —por ejemplo, el generador opera yendo desde lo simple a lo complejo, y se acepta el primer candidato que obtenga éxito.

Una ventaja de esta idea de las limitaciones abiertas radica en su énfasis en que no se puede limitar artificialmente el contex-

to donde se tienen que solucionar los problemas. Todos los problemas dependen no sólo del enunciado específico del problema, sino también de un ámbito de información básica, que se supone debe tener en su cabeza (por así decirlo) el solucionador de problemas. Esta información de fondo es tan amplia que no podría ser suministrada como parte del enunciado del problema. De intentarlo, aunque fuese para humanos, cambiaría drásticamente la naturaleza del problema. ¡Imagínese, por ejemplo, un problema de inducción que fuese acompañado de un folleto de unas 300 páginas con información básica! De la misma manera, en el estado actual de la tecnología, no hay forma de determinar toda la información existente en la cabeza de los solucionadores humanos de problemas.

Una seria deficiencia de la idea de las limitaciones abiertas reside en la carencia de cualquier demostración, salvo por algún ejemplo casual, de que, de hecho, pueda descubrirse así cualquier campo importante de problemas mal estructurados.

2. Métodos débiles. Una segunda hipótesis sobre los problemas mal estructurados proviene de la concepción, descrita anteriormente, de un solucionador general de problemas que posea una secuencia de métodos que sean débiles pero de aplicación amplia. No existe diferencia entre problemas bien y mal estructurados —excepto que un solucionador de problemas llamará mal estructurado a un problema si sólo cuenta con métodos muy débiles para tratar de solucionarlo. La designación depende del solucionador individual de problemas, al igual que del conjunto de métodos utilizables. Pero no existe ninguna evidencia de que nuestra distinción entre problemas bien y mal estructurados sea de hecho muy nítida.

El núcleo operativo de esta hipótesis radica en que no existe ninguna característica especial en las tareas a las que se haya impuesto la etiqueta de “mal estructuradas” que provoque una clase especial de métodos para enfrentarse con tales características especiales. Por el contrario, si se continúa simplemente cons-

truyendo solucionadores de problemas, con conjuntos más y más amplios de métodos débiles, entonces estos solucionadores terminarán siendo capaces de manipular tareas a las que se haya llamado mal estructuradas.

3. *Especificación del problema.* EASTMAN (1969), al trabajar sobre problemas de diseño arquitectónico, lanzó la hipótesis de que los problemas están mal estructurados en la medida en que la actividad de solucionar un problema transcurre al mismo tiempo que la actividad de formularlo. Ha procedido de un modo casi clásico —recogiendo una clase de tareas que parecen sustancialmente mal estructuradas en algunos aspectos, y considerando después con detalle cómo se progresa en su solución. En estos estudios se dedicó a observar la conducta de diseñadores humanos utilizando métodos de protocolo.

La tarea que plantea consiste en reelaborar el diseño de un cuarto de baño en una casa pequeña, proporcionando algunas observaciones negativas para las expectativas de compra en la versión inicial de ese cuarto de baño, más un límite de coste adicional (50 \$). Las observaciones son bastante generales:

Este modelo de casa no se vendió bien. El personal de venta escuchó algunos comentarios de posibles compradores sobre el pobre diseño del cuarto de baño. Se recuerdan varios comentarios: “Qué cantidad de espacio desperdiciado”; “Esperaba encontrar un baño más lujoso”. (EASTMAN, 1969).

No se determina nada sobre los materiales que pueden utilizarse, las condiciones que limiten la iniciativa del diseño, o si los comentarios deben ser realmente tomados con seriedad, etcétera. Los diseñadores son profesionales (en realidad, estudiantes de una escuela de arquitectura que están aprendiendo a ser profesionales). Aportan ya a la tarea actitudes hacia lo que significa ser un arquitecto, al igual que un conocimiento sobre arquitectura y familiaridad con los cuartos de baño.

Dado un juicio inicial, forzosamente informal, de que la tarea está mal estructurada, la forma en que actúan los soluciona-

dores del problema proporciona algunas nociones sobre la mala estructuración.

Todos los solucionadores del problema parten del mismo marco global. El sistema de ejecución consistió en última instancia en un conjunto de limitaciones, un conjunto de objetos (unidades de diseño, tales como bañeras, el espacio libre frente al lavabo, etc.), un conjunto de generadores para crear un diseño, y sistemas de evaluación para decidir si un diseño determinado satisface las condiciones.

Para nosotros, el aspecto interesante del trabajo no radica en la estructura de los programas de los sujetos, sino en que el acto de la creación de unidades de diseño y de limitaciones continuaba a todo lo largo del período de solucionamiento, combinándose con los intentos para solucionar el problema del diseño dentro de la definición de dicho problema de que se dispone en ese momento. Así, el problema está mal estructurado desde el principio para el solucionador de problemas, puesto que muchos (¿la mayoría?) de los aspectos no están definidos en el momento inicial, y la estructuración del problema se desarrolla de un modo continuo.

Estos conceptos no son inconsistentes con la idea de limitaciones abiertas, tal como la propuso Reitman. Pero se subrayan cosas distintas. La concepción de limitaciones abiertas pone de manifiesto el carácter de dependencia contextual del modo en que se cierran las limitaciones. La proposición de Eastman no necesita tan amplia dependencia contextual. El solucionador del problema no conoce algunas de las condiciones y es un problema el hallarlas, lo cual le fuerza a una búsqueda en su memoria y a una evaluación de lo evocado. Pero el solucionador del problema no se compromete de una sola vez, al comienzo de la sesión, en este acto de especificación, para ser luego seguido de la adecuada solución del problema. Por el contrario, tanto la solución del problema como la definición del mismo progresan conjuntamente. De este modo, el problema retiene su carácter de mal definido hasta que se alcanza la solución final.

Pero no hay por qué empeñarse en contrastar estas tres hipótesis como alternativas excluyentes. Cada una genera ciertas ideas sobre lo que pueda ser la solución de problemas con tareas mal estructuradas, y éstas pueden ayudar a encontrar un solucionador de problemas que funcione en tareas peor estructuradas que cualquiera de las tratadas hasta la fecha.

Al comienzo de esta sección (pág. 96 s.), sugerí una serie de tareas generadas bajo (según parecía) tres condiciones: que fuesen problemas mal estructurados; que fuesen tan diversas como fuese posible; y que todavía no se hubiese conseguido mecanizarlas con éxito. Aunque esto último es realmente cierto, no se debe llegar a la conclusión de que no se sabe nada sobre los mecanismos que actúan en tales tareas. Ya hemos comentado la naturaleza del diseño en la discusión del trabajo de Eastman. Se saben más cosas, naturalmente, sobre los procesos de diseño tanto desde el punto de vista de la ingeniería como desde el punto de vista del diseño mecanizado o desarrollado por computadores, aunque una y otra perspectivas tienden a centrarse en aspectos menos relacionados con el carácter mal estructurado del proceso de diseño.

Además, se ha realizado un detallado estudio de las elecciones de carrera por estudiantes (SOLBERG, 1968). Se ha comenzado también a relacionar los fenómenos de descubrimiento científico con los mecanismos de solución de problemas discutidos en la primera parte del presente trabajo (SIMON, 1966).

Ninguna de estas varias parcelas de investigación ha conseguido identificar aspectos de la solución de problemas que sean específicos de la mala estructuración. Quizá no debería pensarse en la noción de mala estructuración como una entidad conceptual única, sino como algo de lo cual vamos separando capas o estratos a medida que vamos avanzando, gradualmente, en el descubrimiento de la naturaleza de la solución de problemas.

CONCLUSION

Después de considerar tres concepciones de la inteligencia artificial y un sinnúmero de problemas, hemos llegado al final. He tratado de presentar el material de la inteligencia artificial en una cantidad suficientemente sustancial para hacer posible una discusión razonable sobre su naturaleza. Me pareció necesario optar por la diversidad, proporcionando tres concepciones distintas del campo, para evitar interpretaciones simplistas.

De hecho, hemos desplegado un vasto abanico de cuestiones. Algunas intentaron captar el interés filosófico por la naturaleza general de la mente y sus relaciones con el computador, ahora que éste se ha instalado entre nosotros. Esta empresa seguramente no tuvo mucho éxito. La inteligencia artificial o bien es irrelevante para los problemas planteados o bien plantea respuestas en forma inaceptable. La principal de ellas fue la propuesta de que la mente es simplemente un sistema de procesamiento de información. Una segunda, aunque forzosamente analógica, fue la propuesta de que la neurofisiología es casi totalmente irrelevante para los problemas de la mente.

Después planteé dos cuestiones, suscitadas en el seno de la concepción de procesamiento de información: sobre la generalidad y sobre problemas mal estructurados. A este respecto expuse una serie de opciones o propuestas surgidas en el campo de la inteligencia artificial, con la esperanza de que sirviesen de ayuda para estructurar la discusión. No se ha ofrecido aquí ninguna opinión final sobre estos problemas, porque de hecho no la tengo.

Estas dos cuestiones se plantearon desde la concepción de la inteligencia artificial como exploración. No son problemas en los que uno se compromete inmediatamente, cuando se interesa

por el modelamiento de la conducta humana. Sin embargo, tienen cierto interés, puesto que son dos de las cuestiones que surgen más a menudo al considerar las limitaciones de los sistemas actuales de inteligencia artificial.

Un tema explícito se ha mantenido a lo largo de todo este ensayo: que los sistemas de procesamiento de información constituyen una “concepción” y no un conjunto de hechos casuales que se aplican o no cuando los problemas locales lo permiten. Dos cosas debieran decirse sobre este aserto. En primer lugar, la referida concepción es de procesamiento de información, no de inteligencia artificial. Pienso que no hay nada de especial en inteligencia artificial que no pueda ser asimilado en la principal corriente de la ciencia de los computadores. En segundo lugar, no se otorga ningún crédito especial a la palabra “concepción”. Algunos de los peores cánceres intelectuales que ha padecido la humanidad se han presentado en forma de “concepciones”. Por una concepción sólo entiendo que se ofrece un marco lo suficientemente amplio para estimular y permitir el pensamiento, de forma que afecte no sólo a las respuestas, sino también a las preguntas y a los criterios de evidencia y relevancia. Es bien sabido que cuando se intenta la comunicación entre defensores de distintas concepciones, lo normal es que sean muy pocas las cosas de valor científico que puedan resultar.

Quizá he debido actuar de manera distinta —como a menudo se nos aconseja hacer en empresas interdisciplinarias—, confiando acaso en que un poco de charla y de fervoroso entusiasmo servirían para establecer una base de comunicación sobre problemas fundamentales. Me pareció más adecuado optar por una aproximación más directa.

NOTA. Este trabajo se desarrolló originalmente para una conferencia sobre el Concepto de Mente, por el Grupo de Estudio de la Unidad de Conocimiento de Berkeley, California, en Agosto de 1969. La audiencia original tenía primariamente un interés filosófico, y algo, aunque no todo, de esa intención se ha conservado dentro de la actual revisión.

Quisiera agradecer a Richard Young sus comentarios sobre el proyecto

primitivo. Como siempre, mi deuda con Herbert Simon está más allá del simple agradecimiento. Muchas de las actitudes defendidas en este ensayo encuentran expresión alternativa en su breve libro sobre *Las ciencias de lo artificial* (SIMON, 1969) y en una reciente empresa conjunta (NEWELL y SIMON, 1972).

El trabajo sobre el que se apoya este escrito fue sufragado en parte por la Advanced Research Projects Agency of the Office of the Secretary of Defense (F 446220-67-C-0058), con el asesoramiento de las Air Force Office of Scientific Research, y en parte por la Research Grant MH-07722-05 del National Institut of Health.

REFERENCIAS BIBLIOGRAFICAS

- ANDERSON, A.R., 1964. "Introduction". En ANDERSON, ed., *Minds and Machines*. Englewood Cliffs, N.J.: Prentice Hall.
- ANSCOMBE, G.E.M., 1958. "Pretending", *Proceedings of the Aristotelian Society*. Supplementary Volume XXXII.
- ATKINSON, R.C. y SHIFFRIN, 1968. "Human Memory: A Proposed System". En SPENCE y SPENCE, eds., *The Psychology of Learning and Motivation: Advanced in Research and Theory*. Vol. 2, Nueva York: Academic Press.
- BELOFF, J., 1967. "The Identity Hypothesis: A Critique". En SMYTHIES, ed., *Brain and Mind*. Londres: Routledge & Kegan Paul.
- BOWER, G.H., 1967. "A Multicomponent Theory of the Memory Trace". En SPENCE, ed., *The Psychology of Learning and Motivation: Advances in Research and Theory*. Vol. 1. Nueva York: Academic Press.
- BROADBENT, D.E., 1958. *Perception and Communication*. Londres: Pergamon Press.
- BRUNER, J.S., GOODNOW, J.J. y AUSTIN, G.A., 1956. *A Study of Thinking*. Nueva York: Wiley.
- CHASE, W.G. y SIMON, H.A., 1973. "Perception in Chess". *Cognitive Psychologie*, 4 (1), 55-81.
- CHOMSKY, N., 1965. *Aspectos de la teoría de la sintaxis*, 1965. Introducción, versión, notas y apéndice de C.P. Otero. Madrid: Aguilar, 1970.
- CLARK, H.H., 1969. "Linguistic Processes in Deductive Reasoning", *Psychological Review*, 75, 4, 387-404.
- COLBY, K.M. y SMITH, D.C., 1969. "Dialogues between Humans and an Artificial Belief System". En WALKER y NORTON.
- COLBY, K.M., TESLER, L. y ENEA, H., 1969. "Experiments with a Search Algorithm on the Data Base of a Human Belief Structure". En WALKER y NORTON.
- COLES, L.S., 1969. "Talking with a Robot in English". En WALKER y NORTON.
- DIGGORY, S., ed., 1972. *Information Processing in Children*. Nueva York: Academic Press.
- DUCASSE, C.M., 1965. "Minds, Matter and Bodies". En J.R. SMYTHIES, ed., *Brain and Mind*. Londres: Routledge & Kegan Paul.
- EASTMAN, C., 1969. "Cognitive Processes and Ill-Defined Problem". En WALKER y NORTON.
- ERNST, G y NEWELL, A., 1969. *GPS: A Case Study in Generality and Problem Solving*. Nueva York: Academic Press.

- EVANS, T., 1964. "An ALGOL Compiler". *Annual Review of Automatic Programming*, Vol. 4, 87-124.
- EVANS, T., 1964. "A Program for the Solution of Geometric-Analogy Intelligence Questions". En MINSKY.
- FEIGENBAUM, E.A. y FELDMAN, J., eds., 1963. *Computers and Thought*. Nueva York: McGraw-Hill.
- FEIGL, H., 1963. "The 'Mental' and the 'Physical' ". En FEIGL, SCRIVEN y MAXWELL, eds., *Concepts, Theories and the Mind-Body Problem*. Minnesota Studies in the Philosophy of Science, University of Minnesota, 370-497.
- FELDMAN, J.A., 1972. "Some Decidability Results in Grammatical Inference and Complexity". En *Information and Control*. Nueva York: Academic Press.
- FELDMAN, J.A., PINGLE, K., BINFORD, T., FALK, G., KAY, A., PAUL, R., SPROULL, R. y TENNENBAUM, J., 1971. "The Use of Vision and Manipulation to solve the 'Instant Insanity' Puzzle". *Prereadings Second International Joint Conference on Artificial Intelligence*. Londres: British Computer Society.
- FEYERABEND, P.K. y MAXWELL, G., eds., 1966. *Mind, Matter and Method*. Minneapolis: University of Minnesota Press.
- FIKES, R., 1969. "A Heuristic Program for Solving Problems Stated as Non-Deterministic Procedures". Ph.D. Thesis, Carnegie-Mellon University. Inédito.
- FLOYD, R., 1967. "Assigning Meanings to Programs". *Proc. Symp. Applied Mathematics*. Amer. Math. Soc., 19, 19-32.
- FODOR, J.A., 1968. *Psychological Explanation*. Nueva York: Random House.
- GREEN, C., 1969. "Application of Theorem Proving to Problem Solving". *Proceedings of the International Joint Conference on Artificial Intelligence*, Bedford, Mass., MITRE Corp. En WALKER y NORTON.
- HABER, R.N., 1969. *Information Processing Approaches to Visual Perception*. Nueva York: Holt, Rinehart and Winston.
- HARARY, F., NORMAN, R.Z. y CARTWRIGHT, D., 1965. *Structural Models: An Introduction to the Theory of Directed Graphs*. Nueva York: Wiley.
- HEMPEL, C.G., 1966. "Recent Problems of Induction". En R.G. COLONY, ed., *Mind and Cosmos*. Pittsburgh: Univ. of Pittsburgh Press.
- HEWITT, C., 1971, "Procedural Embedding of Knowledge in PLANNER", *Proceedings of the Second Joint Conference on Artificial Intelligence*, 167-182. Londres: British Computer Society.
- HUNT, E.B., MARIN, J. y STONE, P.J. *Experiments in Induction*. Nueva York: Academic Press.
- HUTTENLOCHER, J., 1968. "Constructing Spatial Images: A Strategy in Reasoning". *Psychol. Rev.*, 75, 550-560.
- JOHNSON, E.S., 1964. "An Information Processing Model of One Kind of Problem Solving". *Psychol. Monographs*. Whole No. 581.
- KUHN, Ths., 1962. *La estructura de las revoluciones científicas* (1962). Trad. de Agustín Contín. México: Fondo de Cultura Económica 1971.
- LAUGHREY, K., 1969. "Computer Simulation of Short Term Memory. A Component Decay Model". En SPENCE y BOWER, eds., *Psychology of Learning and Motivation*, Vol. 3. Nueva York: Academic Press.
- LENNEBERG, E.H., ed., 1964. *New Directions in the Study of Language*. Cambridge: The MIT Press.
- MACKAY, D., 1965. "From Mechanism to Mind". En J.R. SMYTHIES, ed., *Brain and Mind*. Londres: Routledge & Kegan Paul, 186.
- MANNA, Z., 1969. "Properties of Programs and the First-Order Predicate Calculus". *J. ACM*, 16, 2, 244-255.
- MANNA, Z. y VUILLEMIN, J., 1972. "Five Point Approach to the Theory of Computation". *J. ACM*, 15, 528-536.
- MAXWELL, G. y FEYERABEND, P.K., eds., 1966. *Mind, Matter and Method*. Minneapolis: Univ. of Minnesota Press.
- MILLER, G.A., 1956. "The Magical Number Seven, Plus or Minus Two". *Psych. Rev.*, 63, 81-97.
- MILLER, G.A., 1962. "Some Psychological Studies of Grammar". *American Psychologist*, 17, November, 748-762.
- MINSKY, M., 1961. "Steps toward Artificial Intelligence". *Proc. IEEE*. Reimpreso en FEIGENBAUM y FELDMAN.
- MINSKY, M., 1968. *Semantic Information Processing*. Cambridge: The MIT Press.
- MINSKY, M. y PAPERT, S., 1969. *Perceptrons*. Cambridge: The MIT Press.
- MOSES, J., 1968. "Symbolic Integration". Ph.D. Thesis no publicada. MIT.
- NEWELL, A., 1962. "Learning, Generality and Problem Solving". En M. POPPLEWELL, ed., *Proceed. IFIPS Congress*.
- NEWELL, A., 1967. *Studies in Problem Solving: Subject 3 on the Arithmetic Task DONALD + GERALD = ROBERT*. Pittsburgh: Carnegie Institute of Technology.
- NEWELL, A., 1969. "Heuristic Programming: Ill-Structured Problems". En ARONOFSKY, ed., *Progress in Operations Research*, Vol. III, 360-414. Nueva York: Wiley.

- NEWELL, A., 1970. "Remarks on the relationship between Artificial Intelligence and Cognitive Psychology". En BANERJI y MESARWIC, eds., *Theoretical Approaches to Non-Numerical Problem Solving*. Berlín: Springer Verlag.
- NEWELL, A. y ERNST, G., 1965. "The Search for Generality". En KALENICH, ed., *Proc. of IFIP Congress*, 17-24. Nueva York: Spartan Books.
- NEWELL, A. y SIMON, H., 1961. "Simulación del pensamiento humano", (1961). *Teorema IV/3*, (1974), 335-78.
- NEWELL, A. y SIMON, H., 1963. "GPS: A Program that simulates Human Thought". En FEIGENBAUM y FELDMAN.
- NEWELL, A. y SIMON, H., 1965. "An Example of Human Chess Play in the Light of Chess Playing Programs". En WEINER y SCHADE, eds., *Progress in Biocybernetics*, Vol. 2, 19-75. Nueva York: American Elsevier.
- NEWELL, A. y SIMON, H., 1967. "Overview: Memory and Process in Concept Formation". En KLEIMUNTZ, ed., *Concepts and the Structure of Memory*, 241-262. Nueva York: Wiley.
- NEWELL, A. y SIMON, H., 1972. *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall.
- NILSSON, N., 1971. *Problem Solving Methods in Artificial Intelligence*. Nueva York: McGraw-Hill.
- NORMAN, D.A., 1969. *Memory and Attention*. Nueva York: Wiley.
- NORMAN, D.A. y RUMELHART, D.E., 1970. "A System for Perception and Memory". En NORMAN, ed., *Models of Human Memory*. Academic Press.
- PAIVIO, A., 1971. *Imagery and Verbal Processes*. Nueva York: Holt, Rinehart and Winston.
- PUTNAM, H., 1960. *Mentes y máquinas* (1960). En preparación en la serie "Cuadernos Teorema".
- QUINTON, A., 1965. "Mind and Matter". En J.R. SMYTHIES, ed., *Brain and Mind*. Londres: Routledge & Kegan Paul.
- REITMAN, W., 1965. *Cognition and Thought*. Nueva York: Wiley.
- ROBINSON, J.A., 1967. "Un panorama de la prueba automática de teoremas" (1967). En preparación en la serie "Cuadernos Teorema".
- SCRIVEN, M., 1966. "The Limitations of the Identity Theory". En FEYERABEND y MAXWELL.
- SCRIVEN, M., 1972. "The Concept of Comprehension". En CARROLL y FREEDLE, eds., *Language Comprehension*. Washington: J.V. Winston & Co.
- SIMON, H.A., 1966. "Scientific Discovery and the Psychology of Problem Solving". En COLODNY, *Mind and Cosmos*. Pittsburgh: Univ. of Pittsburgh Press.
- SIMON, H.A., 1969. *Las ciencias de lo artificial* (1969). Barcelona: A.T.E. 1973.
- SIMON, H.A. y FEIGENBAUM, E.A., 1964. "An Information-Processing Theory of Some Effects of Similarity, Familiarization and Meaningful in Verbal Learning". *J. Verb. and Verb. Behav.*, 3, 385-396.
- SIMON, H.A. y KOTOVSKY, K., 1963. "Human Acquisition of Concepts for Sequential Patterns". *Psych. Rev.*, 70, 534-546.
- SLAGLE, J., 1965. "A Multipurpose Theorem Proving Heuristic Program that Learns". *Proc. IFIP Congress 65*. Vol. 2, 323-324. Nueva York: Spartan Books.
- SLAGLE, J.R., 1971. *Artificial Intelligence: The Heuristic Programming Approach*. Nueva York: McGraw-Hill.
- SOLBERG, A., 1968. "A Study of Decision Making: Job Choice". Ph.D. Thesis, Carnegie-Mellon University.
- SPERLING, G., 1960. "The Information Available in Brief Visual Presentation". *Psychol. Monographs*, 74, Whole No. 498.
- TURING, A.M., 1950. *¿Puede pensar una máquina?* (1950). "Cuadernos Teorema" núm. 1, Valencia, 1974.
- WALDINGER, R.J. y LEE, R.C.T., 1969. "PROW: A Step Toward Automatic Program Writing". En WALKER y NORTON.
- WALKER, D.E. y NORTON, L.M., 1969. *Proceedings of the International Joint Conference on Artificial Intelligence*. Boston: Mitre Corporation.
- WANG, H., 1960. "Hacia una matemática mecánica" (1960). En preparación en la serie "Cuadernos Teorema".
- WATERMAN, D. y NEWELL, A., 1971. "Protocol Analysis as a Task for Artificial Intelligence". *Artificial Intelligence* 2, 285-318.
- WAUGH, N.C. y NORMAN, D.A., 1965. "Primary Memory". *Psychol. Rev.*, 72, 89-104.
- WHITEHEAD, A.N. y RUSSELL, B., 1910. *Principia Mathematica*. Cambridge: Cambridge Univ. Press.
- WILLIAMS, D., 1969. "Computer Program Organization Induced by Problem Examples". Ph.D. Thesis, Carnegie-Mellon Univ. Inédito.
- WINIKOFF, A., 1967. "Eye Movements and an Aid to Protocol Analysis of Problem Solving Behavior". Ph.D. Thesis, Carnegie-Mellon Univ. Inédito.
- WINOGRAD, T., 1972. *Understanding Natural Language*. Nueva York: Academic Press.
- ZIFF, P., 1959. "The Feelings of Robots". *Analysis*, XIX, 3. Reimpreso en A.R. ANDERSON.

Cuadernos Teorema

- 1/A. M. Turing
¿PUEDE PENSAR UNA MÁQUINA?
- 2/A. J. Ayer, E. Gellner, I. V. Kuznetsov
FILOSOFÍA Y CIENCIA
- 3/J. Łukasiewicz
PARA UNA HISTORIA DE LA LÓGICA DE ENUNCIADOS
- 4/E. W. Beth
LAS PARADOJAS DE LA LÓGICA
- 5/M. O. Beckner
EL DARWINISMO
- 6/N. Chomsky, M. Foucault
LA NATURALEZA HUMANA: ¿JUSTICIA O PODER?
- 7/P. K. Feyerabend
CÓMO SER UN BUEN EMPIRISTA
- 8/K. Gödel
SOBRE PROPOSICIONES FORMALMENTE INDECIDIBLES
DE LOS PRINCIPIA MATHEMATICA Y SISTEMAS AFINES
- 9/L. Kolakowski, St. Hampshire
EL MITO DE LA AUTOIDENTIDAD HUMANA.
UNIDAD DE LA SOCIEDAD CIVIL Y LA SOCIEDAD POLÍTICA
- 10/N. R. Hanson
EN LO QUE NO CREO
- 11/D. Hume
UN COMPENDIO DE UN TRATADO DE LA NATURALEZA
HUMANA
- 12/J. M. Bochenski
LÓGICA Y ONTOLOGÍA
- 13/J. Habermas
LA CRÍTICA NIHILISTA DEL CONOCIMIENTO EN NIETZSCHE
- 14/N. Chomsky
CUESTIONES DE FORMA Y DE INTERPRETACIÓN
- 15/J. R. Searle
¿QUÉ ES UN ACTO DE HABLA?
- 16/Aristóteles
PERI HERMENEIAS. DE INTERPRETATIONE
- 17/H. Albert
ÉTICA Y METAÉTICA

- 18/E. W. Beth
ENTRAÑAMIENTO SEMÁNTICO Y DERIVABILIDAD FORMAL
- 19/C. Lévi-Strauss
CRITERIOS CIENTÍFICOS EN LAS DISCIPLINAS SOCIALES Y HUMANAS
- 20/J. D. Gergely Bacca
FILOSOFÍA Y TEORÍA DE LA RELATIVIDAD
- 21/L. Tolstol
SOBRE ARTE
- 22/J. Stuart-Mill
DE LOS CUATRO MÉTODOS DE INDAGACIÓN EXPERIMENTAL
- 23/P. Lorenzen
PENSAMIENTO METÓDICO
- 25/G. Frege
BEGRIFFSSCHRIFT. CONCEPTOGRAFÍA
- 26/St. Stich, N. Chomsky, J. J. Katz
DEBATE SOBRE LA TEORÍA DE LA CIENCIA LINGÜÍSTICA
- 27/A. Newell
INTELIGENCIA ARTIFICIAL Y CONCEPTO DE MENTE
- 28/G. H. von Wright
LÓGICA DEÓNTICA
- 30/B. Russell, F. C. Copleston
DEBATE SOBRE LA EXISTENCIA DE DIOS
- 31/A. E. Musgrave
LOS SEGUNDOS PENSAMIENTOS DE KUHN
- 34/M. Dufrenne
ARTE Y LENGUAJE
- 35/H. Skolimowski
RACIONALIDAD EVOLUTIVA
- 36/F. Nietzsche, H. Vaihinger
VERDAD Y MENTIRA EN SENTIDO EXTRAMORAL.
LA VOLUNTAD DE ILUSIÓN EN NIETZSCHE
- 37/Th. S. Kuhn
LA FUNCIÓN DEL DOGMA EN LA INVESTIGACIÓN CIENTÍFICA
- 43/E. Kurzweil
MICHEL FOUCAULT. Acabar la era del hombre
- 51/J. Mosterin
UN CÁLCULO DEDUCTIVO PARA LA LÓGICA DE SEGUNDO ORDEN
- 53/P. Feyerabend, A. Naess
EL MITO DE LA "CIENCIA" Y SU PAPEL EN LA SOCIEDAD
¿POR QUÉ NO CIENCIA TAMBIÉN PARA ANARQUISTAS?

VOLOMENES EN PRENSA

- 24/J. Habermas, Y. Bar-Hillel
TEORÍA DE LA COMUNICACIÓN: DISTORSIÓN Y COMPETENCIA
LA FILOSOFÍA DEL LENGUAJE DE HABERMAS
- 29/H. Feigl, St. Toulmin
EL LEGADO DEL POSITIVISMO LÓGICO
- 32/G. Ryle
CATEGORÍAS
- 39/J. D. Sneed, Th. S. Kuhn
PROBLEMAS FILOSÓFICOS EN LA CIENCIA EMPÍRICA
DE LA CIENCIA
EL CAMBIO DE TEORÍA COMO CAMBIO DE ESTRUCTURA
- 40/Konrad Lorenz
LA DOCTRINA KANTIANA DEL A PRIORI A LA LUZ DE LA
BIOLOGÍA CONTEMPORÁNEA
- 41/D. Hume
LA NORMA DEL GUSTO Y OTROS ENSAYOS
- 42/E. Laszlo, L. von Bertalanffy
HACIA UNA FILOSOFÍA DE SISTEMAS
- 44/H. J. Eysenck
TERAPIA DE CONDUCTA
- 45/R. Descartes
OBSERVACIONES SOBRE LA EXPLICACIÓN DE LA MENTE HUMANA
- 47/H. G. Gadamer
LA DIALÉCTICA DE LA AUTOCONCIENCIA EN HEGEL
- 49/D. Davidson
VERDAD Y SIGNIFICADO
- 55/G. Gentzen
INVESTIGACIONES SOBRE LA DEDUCCIÓN LÓGICA

Información y Distribución: **REVISTA TEOREMA**

Apartado 6060

Valencia-11

A. M. TURING

¿Puede pensar una máquina?

DEPARTAMENTO DE LOGICA Y FILOSOFIA DE LA CIENCIA
UNIVERSIDAD DE VALENCIA

1974

PETER H. LINDSAY
DONALD A. NORMAN

PROCESAMIENTO de INFORMACION HUMANA

Una introducción a la Psicología

*

PERCEPCIÓN
y RECONOCIMIENTO
DE FORMAS

* *

MEMORIA
y
LENGUAJE

* * *

APRENDIZAJE,
CONOCIMIENTO Y DECISIÓN

La obra de Lindsay y Norman es una de las primeras introducciones, prácticamente la única en versión castellana, que distinguiendo entre control de la conducta y mecanismos explicativos de la misma, se decide a incorporar los nuevos enfoques cognitivos, la tecnología del computador, los paradigmas lingüísticos y, en fin, la Psicología como procesamiento de información, para presentar la nueva visión de esta disciplina que ya se practica en múltiples centros internacionales de investigación.

EDITORIAL TECNOS

MADRID
1975 - 1977

JOSEPH WEIZENBAUM

LA FRONTERA ENTRE EL ORDENADOR Y LA MENTE

Los instrumentos
De dónde proviene la fuerza de los
ordenadores
Cómo trabajan los ordenadores
La ciencia y el programador com-
pulsivo
Teorías y modelos
Los modelos con ordenador en la
psicología
El ordenador y el lenguaje natural . .
Inteligencia artificial
Programas incomprensibles
Contra el imperialismo de la razón
instrumental

MADRID 1978