

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.