Allen Newell, *Unified Theories of Cognition.* Cambridge, Mass.: Harvard University Press, 1990.

FERNANDO MARTÍNEZ[†] y JESÚS EZQUERRO[‡]

Departamento de Lógica y Filosofía de la Ciencia
Universidad del País Vasco/Euskal Herriko Unibertsitatea,
Apartado 1249,
20080 SAN SEBASTIÁN,
ESPAÑA

La psicología es una disciplina que se ha caracterizado tradicionalmente por la disgregación de sus conocimientos en multitud de microteorías de corto alcance y de limitado poder explicativo, no tiene por tanto, nada de extraño, que la ciencia cognitiva haya heredado esa pauta metodológica. En el terreno de la cognición, se cuenta con un número considerable de regularidades respecto a un sinfín de tareas experimentales diferentes y, a pesar de que se han ido asentando algunas concepciones generales en áreas como la memoria o el aprendizaje, aún resta mucho para decir que hemos alcanzado una visión medianamente unitaria de todo el aparato cognitivo. De hecho, el esfuerzo investigador en ciencia cognitiva se ha venido concentrando en el diseño de modelos computacionales con los que se intenta simular una u otra habilidad cognitiva. No obstante, quizá por temor a producir frankensteins, se ha venido relegando para mejores momentos

---

[†] Becario de Investigación del Gobierno Vasco.

[‡] Agradecemos el apoyo del Proyecto de Investigación PS92-0041 del Ministerio de Educación y Ciencia (DGICYT), en cuyo marco ha sido redactado este comentario.

la difícil empresa de diseñar teorías unificadoras que no sean resultado de la mera yuxtaposición de las partes[1]. Pero no es ese el único obstáculo, sino que es sólo un síntoma de algo más profundo. A decir verdad, no existe un mínimo acuerdo sobre cómo ha de abordarse la unificación y quién debe realizarla, ni siquiera acerca de si son posibles, o convenientes, las teorías unificadas[2]. Diríase a primera vista que se trata tan sólo de una disputa gremial, de modo que los neurocientíficos arriman el ascua a su sardina aduciendo que es cosa de su exclusiva competencia, y otro tanto dicen los psicólogos; los filósofos, por su parte, argumentan razonablemente que nadie está en mejor posición que ellos para atar ese sin fin de cabos sueltos en que se ha convertido la ciencia cognitiva[3], e incluso los físicos han alegado que es la física elemental la disciplina legitimada para fundamentar, unificar y establecer los límites de las teorías de la cognición[4]. Es esto precisamente lo que se propone Allen Newell: establecer las bases sobre las que se pueda fundar una teoría unificada de la cognición (TUC).

---

[1] Este problema plantea dificultades especiales en caso del diseño de modelos en IA. Cualquier experto en computación sabe de la enorme dificultad que supone tratar de integrar microcomponentes desarrollados independientemente, aun cuando tengan un comportamiento óptimo, en sistemas integrados. Por esa razón, los componentes suelen ser diseñados con vistas a la integración. Otro tanto cabe pensar de las microteorías propuestas independientemente y su posibilidad de integración en teorías comprehensivas.

[2] M. Minsky, por ejemplo, contraargumenta la afirmación de Newell de que toda ciencia progresa tratando de unificar diciendo que mientras ese objetivo parece plausible, y deseable, para la física, no lo es a partir del nivel biológico. Véase M. Minsky (1986) y (1993).

[3] Por ejemplo, D. Dennett admite abiertamente esta interpretación como moraleja de su *Consciousness Explained* (1991)

[4] Véase, si no, el caso de Penrose (1989)

La tesis de Newell se articula en torno a dos supuestos fundamentales: en primer lugar, asume que el grado de desarrollo actual de la psicología es suficiente para plantear tales TUCs, esto es, teorías que postulan un único sistema de mecanismos que operan conjuntamente para producir todo el espectro de la cognición humana; en segundo lugar, mantiene que los posibles candidatos a TUCs deben ser teorías a nivel de mecanismo, esto es, arquitecturas computacionales fijas capaces de procesar contenido variable. Dada la controvertida situación actual acerca del formato básico que deben poseer las teorías de la cognición, la apuesta, se convendrá, no es en modo alguno trivial. Para cumplir estos objetivos, Newell nos ofrece su candidato, una arquitectura que recibe el nombre de Soar, la más desarrollada de las arquitecturas basadas en sistemas de producción, sobre la que una comunidad de expertos en IA, liderada por el propio Newell hasta su muerte, viene trabajando desde hace años. La descripción de las características de Soar se irá desgranando a lo largo de los capítulos y constituye el grueso de la obra. Pero previamente a esto, Newell quiere dejar claro cuál es su concepción de la psicología y de la ciencia cognitiva en general. A ello dedica los tres primeros capítulos de los ocho que constituyen el libro.

Como señala en el primer capítulo, que hace las veces de introducción y justificación de la empresa en la que se va a embarcar, el suyo no es el primer intento de presentar algo parecido a una TUC. Entre los precursores que menciona cabe destacar a Anderson y su Act*, quien también formula su teoría en forma de arquitectura cognitiva. A pesar de las diferencias entre autores o sistemas, Newell se esfuerza por mostrar un campo en progreso con un paradigma firmemente establecido, el que podríamos llamar cognitivismo clásico, en el que la tarea principal por hacer (la que debe acometer una TUC) es relacionar el inmenso caudal de datos. La teorización

psicológica (y científica en general) se concibe como un proceso acumulativo de cara a conseguir teorías cada vez más aproximadas pero siempre modificables. En este proceso Newell establece una lista de prioridades que una TUC debe ir satisfaciendo, comenzando por capacidades como la solución de problemas o la toma de decisiones y terminando con fenómenos como la imaginación o el sueño. Se echa de menos en este punto una justificación de esta jerarquía. Tácitamente se nos hace pensar que debemos empezar por aquellos fenómenos en los que se muestra claramente la inteligencia pero no demasiado complejos como para frustar nuestra investigación desde el principio. Por esa razón se relega el lenguaje a un estadio posterior, detrás incluso de fenómenos como la memoria o la percepción. Debe decirse, no obstante, que esta jerarquía no tiene por qué ser idéntica para cualquier TUC, como parece suponer Newell. Los sistemas conexionistas (otros posibles candidatos para unificar) comienzan, por ejemplo, por tareas de más bajo nivel y encuentran menos complejidad en tareas basadas en el reconocimiento de patrones. El fondo de la cuestión es que Newell, como ya se ha dicho, se inserta de lleno en el punto de vista clásico, cuya propuesta consiste en sistemas simbólicos, mejor pertrechados para las capacidades en primer lugar de la lista (es precisamente por la solución de problemas por donde comenzó junto com Simon y Shaw para desarrollar su *General Problem Solver* a finales de los 50). En el siguiente capítulo va a presentar, por tanto, las bases de la ciencia cognitva desde el punto de vista clásico.

Newell considera que un sistema es inteligente en la medida en que se aproxima a lo que él denomina nivel del conocimiento[5], un nivel que es una abstracción del

[5] Que ya había presentado en su célebre artículo de 1982.

procesamiento y la representación internos. Es el superior de una jerarquía de niveles y puede especificarse en función del conocimiento que posee el sistema y de las metas que persigue. Como él mismo señala, este tratamiento es similar al que se da desde determinados estudios filosóficos de la intencionalidad, especialmente la noción de sistemas intencionales (Dennett (1987)). Este conocimiento debe ser representado de algún modo y, en principio, Newell se muestra bastante neutral respecto al medio representacional siempre y cuando cumpla ciertas características, entre las que destaca la composicionalidad. A pesar de que se mencionen distintas representaciones, como las analógicas, y sus limitaciones, la exposición parece insuficiente de cara al debate actual sobre la naturaleza de las representaciones en ciencia cognitiva. No hay que olvidar que el modo de representación puede convertir en tratables, o bien en intratables, determinados problemas. Por otra parte, es obvio que el tipo de representación va a quedar bastante determinado por la arquitectura elegida, lo que en su caso signifca adoptar algo cercano a la lógica. La composicionalidad la proporcionan los sistemas computacionales que requieren sistemas de símbolos (SS) como medio de acceder a estructuras distales en la memoria. Newell se reafirma así en su hipótesis de los sistemas simbólico-físicos (Newell (1980)) como requisitos para realizar un sistema de conocimiento. El SS se realiza en una arquitectura, noción clave puesto que, como hemos dicho, para Newell una TUC ha de formularse en forma de arquitectura, i.e. la estructura fija que ofrece una descripción del sistema al nivel de registro-transferencia. Su noción de arquitectura consuma la separación entre estructura y contenido (que es variable) criticada por hoy desde algunos medios conexionistas. Para exhibir inteligencia un SS requiere además poseer el tipo de procesamiento adecuado, que será la

búsqueda (*search*). La relación entre la cantidad de búsqueda empleada y la cantidad de respuestas preparadas que posee el sistema nos dará el balance entre deliberación y preparación. Este balance define un espacio en cuyas distintas regiones se sitúan diferentes especies de sistemas inteligentes. De su análisis se deduce que una arquitectura que pretenda dar cuenta de la inteligencia humana deberá tender hacia un mayor grado de preparación y no hacia un mayor volumen de búsqueda.

Las carcterísticas de la arquitectura cognitiva humana son desarrolladas en el capítulo tercero, al que Newell cuelga la etiqueta de "especulativo". En él analiza principalmente la restricción del tiempo real, i.e. la medida temporal de las funciones que realiza la cognición humana de acuerdo a la tecnología en la que está instanciada: el sistema nervioso. Un humano es también un SS, y su arquitectura, al igual que en los sistemas atificiales, está construida de un jerarquía de niveles, al ascender por la cual disminuye la velocidad. Newell distribuye los niveles en cuatro bandas: biológica, cognitiva, racional y social. Partiendo de la base de que la velocidad de los niveles biológicos es conocida empíricamente, predice la velocidad de los superiores, de los cuales el libro se va a centrar en los cognitivos (sólo apunta unas sugerencias muy generales respecto a las dos últimas bandas). La restricción del tiempo real sobre esta banda va a suponer que la arquitectura que se postule deba incorporar algún paralelismo. La banda cognitiva incluye los niveles de acto deliberado, operación simple y tarea-unidad. El análisis de los tiempos predichos para cada nivel que efectúa Newell adolece de cierta vaguedad (aunque hay que recordar el carácter especulativo que se otorga al capítulo) y sólo pretende convencer al lector de su plausibilidad. La vaguedad se acrecienta al alcanzar el nivel de tarea cuyo límite inferior se encuentra en el orden de

los segundos y el superior se extiende hacia la banda racional en el orden de las horas. La división de niveles en esta última banda parece arbitraria y únicamente una conclusión lógica de la progresión temporal inferior. La sospecha de una arbitrariedad similar en la división de niveles cognitivos se vería disipada en el capítulo cinco, al ocuparse de tareas en las que se muestre la arquitectura cognitiva. Sin embargo, queda la duda de que en qué punto las medidas temporales dependen del análisis del experimentador de las diversas tareas que se plantean a los sujetos experimentales (los niveles biológicos tienen, en principio, métodos de medida propios).

Una vez sentada su idea de la cognición Newell procede, en el capítulo cuarto, a describir Soar, su arquitectura candidata a TUC. La infraestructura teórica necessaria para el estudio de la cognición humana la ofrece la Inteligencia Artificial (IA), así que Soar es un sistema realizado en ordenador. En primer lugar se muestran las características de la cognición central en Soar. Es un sistema que utiliza espacios de problemas para formular las tareas y sistemas de producciones para toda la memoria a largo plazo (MLP). Esta segunda es identificada por el autor como la característica más discutible de Soar, i.e. la existencia de una sola estructura de memoria compuesta íntegramente por producciones y con procesos de escritura y lectura únicos. Sobre esto hablará com más detalle en el capítulo 6, al tratar la memoria. Soar emplea espacios de problemas, operadores y objetos con atributos y valores. El lector familiarizado con los sistemas computacionales tradicionales encontrará otras dos peculiaridades principales. Una es el sistema para la resolución de conflictos. Cuando no existe una decisión clara a tomar, el sistema entra en un punto muerto (*impasse*), pero el resultado no es un mensaje de error sino la construcción de una submeta que permite acceder a más conocimiento para salir

del atolladero. El otro aspecto novedoso (y conflictivo) es el uso de un solo medio de aprendizaje: la recodificación (*chunking*) en producciones de los elementos que han conducido a un resultado positivo (la consecusión de una meta), de manera que puedan ser utilizados en situaciones similares posteriores, acelerando el procesamiento y produciendo una generalización implícita. Como se ve, es una forma de ganar en eficacia aumentando el grado de preparación, lo cual, de acuerdo a los planteamientos inicales de Newell supone un acercamiento al modo humano de actuar. Al describir el sistema cognitivo total (que incluye la percepción y la conducta motora) Newell retorna al estilo especulativo para dar una visión general del proceso estímulo-organismo-respuesta. De carácter más empírico es la descripción de R1-Soar, un sistema para probar el funcionamiento de Soar con el conocimiento de un sistema experto, y de Designer-Soar, en la que se examina la capacidad de diseñar algoritmos. El capítulo termina con una argumentación a favor de Soar como sistema inteligente y con algunos apuntes sobre la correspondencia entre los rasgos de Soar y los de la cognición humana.

Sólo le resta a Newell, defensor de siempre del carácter empírico de la investigación en computación (como en Newell & Simon (1976)), probar que Soar es capaz de exhibir conducta inteligente. Los tres capítulos siguientes están esencialmente orientados a esta cuestión. En el quinto se ocupa de la conducta inmediata, i.e. aquellas respuestas a estímulos que se realizan en el orden de los 300 ms a los 3 seg. Es aquí donde se revela la arquitectura, puesto que opera en estos márgenes. En las tareas de este tipo las regularidades descubiertas son miles y la tarea de la TUC (arquitectura) es satisfacer el mayor número posible de ellas. Hay que tener en cuenta, sin embargo, que el ajuste a los datos puede hacerse

en muchas escalas. Las regularidades que es capaz de explicar Soar, en su desarrollo actual, son de orden cualitativo (presencia o no de un efecto) y cuantitativo (valores típicos) pero apenas paramétrico, lo cual tampoco es excesivamente criticable. Soar siempre sigue un mismo esquema de funciones que son necesarias y suficientes para realizar la tarea: Percepción - Codificación - Atención - Comprensión - Plantear Tarea - Intención - Decodificación - Movimiento. De acuerdo a la estimación de los operadores o de los ciclos de producción precisos para ejecutar cada función se pueden realizar predicciones del tiempo que llevará una tarea. Al aplicarlo a la tarea de respuesta simple, de doble elección, de compatibilidad E-R o de mecanografía, Soar es usado como teoría al nivel de operador funcional, y en la tarea de reconocientos de ítems de Sternberg proporciona tanto una explicación cualitativa como a nivel de producción. Se ilustra así cómo es posible incorporar en la arquitectura teorías ya existentes.

Ya se ha señalado que la MLP de Soar está totalmente compuesta por produciones, que sirven para reconecer y recuperar la información. En el capítulo sexto, sobre memoria, aprendizaje y habilidad, Newell explica con más detalle esta noción de una única memoria para el conocimiento semántico, episódico y procedimental. La idea es que la MLP es una unidad funcional, lo cual no supone que deba formar una unidad estructural, las producciones pueden ser estructuralmente heterógeneas. Aun sin entrar en debate, esta idea recortaría un argumento usado por los conexionistas en favor de la homogeneidad estructural que ofrecen sus sistemas. Por otra parte, la memoria a corto plazo (MCP) existe funcionalmente pero no como estructura, y es equivalente a la memoria operativa (*working memory*). Pero en cualquier caso el sistema se puede describir como una

organización clásica de MLP y MCP. La teoría del aprendizaje en Soar es cualitativa. Surge siempre de la actividad orientada a metas en la que hay recodificación, y da cuenta de fenómenos como la transferencia, la repetición o la especificidad de la codificación. El nucleo del aprendizaje es procedimental y a partir de un mecanismo de aprendizaje se pueden construir otros. Newell aprovecha problemas que surgen por el camino, como la recodificación de datos arbitrários, para mostrar cómo pueden obtenerse soluciones, sin modificar la arquitectura, que ofrecen rasgos interesantes tales como el uso del control de búsqueda para guiar los operadores hacia la solución. Soar se usa, una vez más, como muestra de cómo incorporar desarrollos teóricos ya existentes, como la teoría de Epam de aprendizaje verbal o la ley de la práctica para la adquisición de habilidades.

La conducta racional deliberada es el objeto del capítulo séptimo. En este campo el sistema se aproxima al nivel del conocimiento y la conducta es función más de la tarea que de restricciones de arquitectura. Soar es aplicado a tres tipos de problemas: criptoaritmética, silogismos y verificación de oraciones. En ellos es preciso hacer uso de los protocolos de los sujetos experimentales y el sistema debe ser capaz de seguir los movimientos racionales intencionados que efectúan las personas, aunque admitiendo una gran variabilidad de estrategias. Respecto al problema de la racionalidad del teórico que analiza la tarea como posible artefacto experimental, Newell defiende el uso de la simulación como método de control. Resulta especialmente interesante la discusión de la tarea de resolver silogismos, que permite a Newell presentar la discusión entre los partidarios del razonamiento por medio de proposiciones y por modelos mentales. Redefine el problema considerando dos dimensiones representcionales: alcance y coste de

procesamiento. Se define así un continuo en que el sistema proposicional puntuaría alto en ambos mientras que los modelos mentales lo harían bajo. El autor se decanta por una solución intermedia que denomina modelos anotados. Soar admitiría los modelos mentales en cuanto representaciones en un espacio de problemas. No obstante, la asimilación de modelos mentales y espacios de problemas no parece suficientemente argumentada, y hay motivos para pensar que existe un contencioso real.

En el último capítulo Newell retoma un tono más especulativo para investigar lo que él llama las "fronteras" de Soar, es decir, aquello que aún queda fuera de la candidata a TUC. Comienza por el lenguaje, en las vertientes de su comprensión y de su aprendizaje. Respecto a la primera, postula unos operadores de comprensión individualizados para el significado de cada palabra en cada contexto particular. La explosión de contextos a considerar que supondría una propuesta semejante podría quedar controlada si se tiene en cuenta que éstos se limitan a los que aparecieron en el aprendizaje (evidentemente por recodificación), el cual sería un proceso de ir llenando los operadores. Pero permanece el problema de incorporar la gramática y sus regularidades, para lo cual Newell ofrece tan sólo unas sugerencias generales. Explora asimismo la capacidad de Soar para tratar el lenguaje como un fenómeno modular de acuerdo a las características fijadas por Fodor. Es importante señalar que el hecho de que la comprensión del lenguaje quede fuera del alcance de Soar, supone un problema especialmente serio para una propuesta que pretende ser una teoría general de la cognición humana. A este respecto, existen indicios razonables para dudar que incluso futuros desarrollos de una arquiectura basada en producciones, como es Soar, permitan alcanzar ese objetivo. Por una parte, en Soar

todo el conocimiento es accesible de manera uniforme, al contrario de lo que suede con estructuras basadas en redes, como los *scripts* o los *frames*, cosa que le resta plausibilidad psicológica. Si a ello añadimos la neta separación entre estructura y contenido asumida por Soar, la consecuencia puede ser que una capacidad tan esencial como el razonamiento ordinario implicado en la comprensión del lenguaje natural resulte computacionalmente intratable[6]. Un segundo campo en el que se adentra es el del desarrollo, para lo cual utiliza los experimentos de balanzas de Piaget. Dado que Soar es una teoría de la inteligencia ya formada, el problema en este caso es mostrar que sus mecanismos (especialmente de aprendizaje) pueden dar lugar a los tipos de pautas observados en los niños sujetos del experimento. Su afirmación central es que sucesivas aplicaciones de la recodificación pueden dar lugar a la construcción de reglas que están al alcance de sujetos de edades paulatinamente superiores. El tercer campo es la relación de Soar con lo bilógico. Su discusión se centra aquí en comparar las virtudes del conexionismo (como familia de sistemas más cercanos a la tecnología neural) con las de Soar. Recalca algunas similitudes funcionales y desecha dos nociones importantes en esos sistemas: la satisfacción de restricciones múltiples, por ser demasiado potente para ser real, y el nivel subsimbólico, por

---

[6] Algunos autores, como Minsky (1993), han argumentado razonablemente que la tratabilidad computacional del razonamiento ordinario debe pasar por la idea de que el cerebro contiene algunas estructuras de tipo red especialmente preparadas para ello. Contra la tesis de la separación neta entre estructura y contenido, proponiendo, alternativamente, la idea de la existencia de una relación simbiótica entre el tipo de representación y la arquitectura computacional para poder dar cuenta del razonamiento que tiene lugar en la escala de milisegundos, puede verse L. Shastri (1990).

no existir espacio temporal en la jerarquia de niveles. No obstante, dado que Newell admite, como acabamos de decir, que Soar es una arquitectura de inteligencia ya formada, y no un producto de la evolución, se echan de menos algunas palabras acerca de si Soar *podría* ser un producto de la evolución, pues en caso contrario, su plausibilidad biológica quedaría minada de raíz. El campo siguiente es la banda social, donde examina someramente cómo contribuye la psicología individual a la social. Termina el capítulo ofreciendo unas ideas sobre el papel de las aplicaciones en las teorías cognitivas y una lista final de recomendaciones para el desarrollo de TUCs.

La reciente y desafortunada muerte de Newell, despues de 37 años en primera línea del desarrollo de la IA, no debe oscurecer la importancia de esta obra. De hecho, su publicación ya suscitó el suficiente interés como para que le fuera dedicado un volumen completo de *Artificial Intelligence* (el 59, 1993) con numerosos comentarios críticos que el propio Newell tuvo ocasión de contestar pero no de ver finalmente publicados. En conjunto, el libro es una excelente muestra de la visión cognitivista clásica de la psicología, instrumentada en un sistema de IA. Pero no se limita a reproducir nociones tradicionales, sino que las pasa por el tamiz de sus más de tres décadas de investigación y las somete a control experimental en un sistema artificial concreto. Si adolece de vaguedad en algunos momentos es por su vocación generalista, y casi siempre se hace de forma deliberada con el fin de ofrecer una visión lo más amplia posible. Aparte de lo que Soar sea o no capaz de hacer, se pretende mostrar cómo se debe utilizar una arquitectura cualquiera para que sea una TUC. Por otro lado, el vasto conocimiento mostrado por Newell del área que se trata se refleja en la multitud de microteorías y resultados experimentales que se nos van

ofreciendo a largo de los capítulos, los cales constituyen su mejor argumento de que realmente hay tela para hilar. Se hecha en falta, una mayor discusión con otros enfoques capaces de producir sistemas candidatos a TUC, incluido el conexionismo, cuya aparición en las páginas es relativamente breve. En definitiva, en una época en que muchos comienzan a hablar en favor de una unificación "por abajo" de la teorización en la ciencia cognitiva (esto es, la que proporcionarían las neurociencias) Newell se atreve a seguir defendiendo la concepción clásica para intentar una unificación "por arriba". No ignora, sin embargo, las restricciones procedentes de la biología, como la del tiempo real, y su formación en ciencias de la computación le hace estar atento a cuestiones de implementación, como se reclama desde distintos ámbitos, especialmente los afines al conexionismo. Este libro constituye así una propuesta para progresar conjuntamente hacia la comprensión de la cognición humana (se incluye también un e-mail para contactar con la comunidad Soar) y condensa en buena parte el importante legado de Newell para el desarrollo de la ciencia cognitiva.

REFERENCIAS

DENNETT, D.C. (1987). *The Intentional Stance.* Cambridge, Mass.: MIT Press, Bradford Books.

———. (1991). *Consciousness Explained,* Boston, Little, Brown.

MINSKY, M. (1986). *The Society of Mind,* N. York: Simon and Schuster.

———. (1986). Allen Newell, *Unified Theories of Cognition* (Book Review), *Artificial Intelligence* 59, pp. 343-54.

NEWELL, A. (1980). Physical Symbol Systems. *Cognitive Science*, 4: 135-183.

———. (1982). The Knowledge Level. *Artificial Intelligence*, 18: 87-127.

NEWELL, A. & SIMON, H. (1976). Computer Science as Empirical Enquiry. En J. Haugeland (Ed.) *Mind Design.* Montgomery, Vt: Bradford Books, pp. 35-66.

PENROSE, R. (1989). *The Emperor's New Mind*, Oxford University Press.

SHASTRI, L. (1990). Connectionism and the computational effectiveness of reasoning. *Theoretical Linguistics*, 16 (1), 65-87.