*Teorías computacionales de la cognición**

HERBERT A. SIMON
Universidad Carnegie Mellon

A finales de los años cincuenta, fue propuesta la hipótesis de que el pensamiento humano es procesamiento de información, es decir, manipulación de símbolos. Como la mayoría de las nuevas ideas, ésta tuvo muchos precursores, especialmente aquéllos reunidos bajo la rúbrica general de la cibernética. Allen Newell y yo hicimos algunas consideraciones sobre estos precursores en el artículo histórico suplementario de nuestro *Human Problem Solving* (1972). Lo novedoso, que comenzó alrededor de 1956, fue la traducción de esas ideas a los programas de ordenador simbólicos (no-numéricos) que simulaban la actividad mental humana en el nivel simbólico. Los aspectos de estos programas podían ser comparados en algunos detalles con los datos que siguen los caminos reales del pensamiento humano (en particular los protocolos verbales) en una variedad de tareas intelectuales, y la veracidad de los programas como teorías del pensamiento humano podían de ese modo ser puestos a prueba.

Así pues el ordenador digital proporcionó a la vez un modo (los programas) para expresar teorías de la cognición precisas y un medio (la simulación, usando esos programas) para comprobar los grados de correspondencia entre las predicciones de la teoría y un comportamiento humano real. Como ejemplos tempranos destacados que tuvieron éxito en ajustarse considerablemente a los registros del comportamiento humano tenemos a EPAM, un programa que simula la percepción y aprendizaje humanos, y GPS, un programa que simula la resolución de problemas humana. Durante estos años, EPAM ha extendido de forma continuada los registros de tareas que maneja, mientras GPS ha sido transmutado en Soar, un paso fundamental en la unificación de las teorías cognitivas. Ambos están todavía muy activos.

* Traducción de Susana García Pardo

Antes de que la propuesta del procesamiento-de-información pudiera jugar un papel sustancial en psicología, tuvo que vencer un número de creencias profundamente arraigadas, asociadas con el conductismo, por un lado, y con la psicología de la Gestalt, por el otro. Los conductistas recelaban de los intentos de teorizar sobre lo que sucedía dentro de la cabeza (aunque el conductismo más duro había sido ya un tanto suavizado en esta dimensión por las teorías de Tolman y Hull). Los gestaltistas se oponían al reduccionismo y a las explicaciones mecánicas de los fenómenos que ellos consideraban «intuitivos» y «perspicaces». Había también una tercera perspectiva, la cual insistía en que las «verdaderas» explicaciones de los fenómenos psicológicos debían ser fisiológicas y neuronales: no había lugar para un nivel de teorías simbólicas entre el comportamiento y el cerebro biológico. Todos estos puntos de vista tenían que ser reconciliados con el simbólico para que este último ganase credibilidad.

Por último, en esos primeros años, pocos psicólogos tuvieron oportunidades para una interacción práctica con ordenadores. Para la mayoría de ellos, el ordenador era un aparato «mecánico» para el cálculo-numérico, una caja negra ocupada por 0-1 bits. Difícilmente podía considerarse un candidato prometedor para representar la flexibilidad, falibilidad y la riqueza del pensamiento humano.

La idea bastante novedosa decía que *un programa era análogo a un sistema de ecuaciones diferenciales (o de diferencias), por lo tanto podía expresar una teoría dinámica.* Igualmente novedosa era la idea de que *los ordenadores no estaban reducidos a lo numérico, sino que podían representar símbolos (patrones con denotaciones) de todo tipo.* El pensamiento en voz alta era confundido con la introspección, de ahí que no se considerara como una fuente legítima de datos empíricos. Las técnicas estadísticas clásicas no eran de mucha utilidad a la hora de juzgar la calidad de la correspondencia entre los aspectos del ordenador y los protocolos verbales.

La hipótesis del sistema de símbolos físico

Pasaron varias décadas antes de que las ideas básicas del procesamiento de la información se propagaran por la comunidad psicológica y comenzaran a sustituir a los supuestos consolidados. En los setenta el nuevo planteamiento había conseguido una posición dominante en la psicología cognitiva. Aún así, había muchos más psicólogos que, mientras aceptaban la mayor libertad de experimentación y teorización que acompañaba a la revolución, no estaban muy entusiasmados con la «me-

táfora del ordenador», como era llamada a menudo, o con la validez de los protocolos verbales como datos. La mayoría de las investigaciones en cognición continuaron usando los métodos y diseños experimentales convencionales y la mayoría de las teorías continuaron siendo expresadas en términos verbales poco precisos.

Fue sólo después de que muchos psicólogos comenzaran a interactuar con ordenadores personales, adquiriendo así una imagen más sofisticada de lo que es un ordenador, cuando creció la apreciación de la idea de un programa de ordenador como una teoría, o la idea de comprobar teorías por comparación de las salidas del ordenador con las verbalizaciones humanas. La opinión del procesamiento de información en psicología que yo acabo de esbozar no está extendida en toda la profesión incluso hoy en día. Aquellos de nosotros que contemplamos los programas de ordenador como «teorías» más que como «metáforas» somos probablemente todavía una minoría. «La Inteligencia Artificial *débil*», como es a veces llamada la perspectiva metafórica, probablemente tiene aún más partidarios que la «Inteligencia Artificial *fuerte*».

El supuesto básico que subyace a la perspectiva fuerte es la hipótesis del sistema de símbolos físico, la cual expusimos Allen Newell y yo en nuestra Conferencia del Premio Turing ACM en 1975 (Newell & Simon, 1976):

> *Un sistema de símbolos físico tiene los medios necesarios y suficientes para la acción inteligente general.*

Un *sistema de símbolos físico* (SSF) es un sistema que tiene las capacidades de un ordenador digital moderno: entrada y salida de símbolos, los cuales organiza y reorganiza en estructuras de símbolos, los almacena y los borra, los compara para identificar o diferenciar, y se comporta contingentemente sobre los resultados de tales comparaciones. Los *símbolos* en un SSF son sencillamente patrones de algún tipo (y hechos de algo) que *apuntan* hacia o *denotan* algo distinto de ellos mismos. Debería destacarse que no hay supuesto alguno con respecto a que los símbolos en un SSF sean en algún sentido verbales – pueden ser patrones de *cualquier* tipo: verbales, pictóricos, abstractos, neuronales. En el cerebro humano, tales patrones a veces son y a veces no son accesibles al conocimiento consciente. Pueden denotar otros símbolos en el cerebro u objetos externos o configuraciones.

La hipótesis del SSF reclama dos tesis empíricas:

1. Que un SSF puede ser programado para comportarse inteligentemente,
2. Que los seres humanos son inteligentes en virtud de ser sistemas de símbolos físicos; su comportamiento inteligente tiene que ser explicado en términos de símbolos y procesos simbólicos.

Es fácil ver la conexión entre estas tesis y la versión «fuerte» de la psicología del procesamiento de información: porque los procesos simbólicos que explican la inteligencia humana son también viables para los ordenadores, nosotros podemos escribir programas de ordenador que producen comportamiento inteligente usando procesos que siguen de cerca los procesos de la inteligencia humana. Estos programas, los cuales predicen cada paso sucesivo en el comportamiento como una función del estado actual de las memorias junto con las entradas actuales, son teorías, bastante análogas a los sistemas de ecuaciones diferenciales de las ciencias físicas. Para poner a prueba estas teorías, necesitamos los datos sobre los procesos del pensamiento humano momento-a-momento; los datos más afinados de este tipo actualmente disponibles son los protocolos verbales de pensamiento-en-voz-alta.

Durante las primeras décadas de la psicología del procesamiento de información, los procesos de pensamiento supuestos eran predominantemente procesos *seriales*, uno-cada-vez. En años recientes, ha habido un creciente interés en los sistemas *paralelos* como una arquitectura alternativa. Mi opinión personal es que hay lugar para ambos en el trabajo del cerebro; pero una determinación final de los papeles respectivos del procesamiento serial y paralelo en el pensamiento humano descansa en el futuro, y no discutiré la cuestión en este trabajo.

Acompañando al interés por las arquitecturas paralelas, se ha incrementado la popularidad de los sistemas que intentan incorporar al menos algo de la organización neuronal del cerebro humano. Estos sistemas *conexionistas* renuncian a un nivel de teorización puramente simbólico en favor de la representación de las actividades cognitivas directamente por una red cuyos elementos son de algún modo en su carácter como nervios. Las «neuronas» pueden ser bastante abstractas, como en la mayoría de los sistemas conexionistas (Rumelhart & McClelland, 1986), o pueden intentar capturar lo más posible de las propiedades físico-químicas de las neuronas reales. En cada caso, estos sistemas tienden a volver a una tradición más antigua que no ve la necesidad de un nivel de teorización simbólico sobre el nivel neurológico. Hasta ese momento en el que el conexionista y otros modelos de redes sean capaces de manejar un abanico de tareas cognitivas complejas comparable a aquéllas

ya explicadas por las teorías de nivel-de-símbolos, parece que no hay mucho interés en intentar resolver este asunto.

Hay también algunos desacuerdos en cuanto a si los sistemas conexionistas y las redes neuronales deben ser tratados como simbólicos (como SSF). Habiendo discutido esta cuestión en otra parte (Vera & Simon, 1994), no la retomaré aquí.

El cuerpo de pruebas empíricas en apoyo del SSF como una teoría de la cognición humana (en la primera de las variantes que he descrito) es por ahora muy grande. Uno puede conseguir un cuadro general de ello desde fuentes tales como Newell y Simon (1972), Anderson (1983), Simon (1979, 1989), Langley y otros (1987), y Newell (1990). El éxito en las aplicaciones de la teoría recorre todo el camino desde los experimentos de aprendizaje clásico, a la resolución de problemas, la obtención de conceptos, el aprendizaje desde ejemplos, la comprensión de instrucciones escritas, el aprendizaje del lenguaje natural (Siklossy, 1972), recuerdo libre, jugar al ajedrez, la percepción visual, las imágenes mentales y el descubrimiento científico. Las referencias arriba citadas son sólo una muestra, y excluyen numerosos sistemas inteligentes en IA que no pretenden simular procesos humanos.

A veces encuentro sorprendente (y no algo frustrante) que esta literatura empírica sea rara vez citada, e incluso menos frecuentemente examinada en detalle, en las discusiones sobre la validez de la hipótesis del SSF. Todavía aparece como una creencia generalizada que la naturaleza de los procesos del pensamiento humano puede ser determinada desde los *primeros principios* sin examinar el comportamiento humano con meticuloso detalle y comparando el comportamiento con la demanda de teorías rigurosas. Yo intentaré evitar ese error en el resto de este artículo; cuando haga reivindicaciones empíricas me referiré a la investigación empírica relevante.

Importancia filosófica del SSF

Antes de volver a asuntos más empíricos, me gustaría comentar algo sobre las implicaciones de la hipótesis del sistema de símbolos físico con respecto a algunos tópicos clásicos en filosofía. El primero de ellos es la epistemología. El segundo es el problema mente-cuerpo.

Epistemología para ordenadores.

La epistemología se preocupa por la cuestión de cómo, dado que vivimos, por decirlo de alguna manera, dentro de nuestras cabezas, ad-

quirimos conocimiento de lo que está fuera de nuestras cabezas. El idealismo esquiva la cuestión al situar todo el interés en el interior de la cabeza, de ese modo se salva de cualquier problema de transporte. El empirismo, en cualquiera de sus formas, carece de esta huida. Quine, en *Palabra y Objeto* (1960), abandonó el intento de averiguar cómo *yo* (Quine) conozco, y se preguntó en su lugar cómo *él* (un informante nativo) conoce. Entonces todo quedaba fuera de la cabeza – al menos de la cabeza de Quine – y las dificultades de tratar con la sensación y la percepción no tenían que ser enfrentadas. Ahora bien, eran sustituidas por la dificultad de que el interior de la cabeza del informante nativo era casi tan inaccesible a Quine como el mundo exterior a su propia cabeza (la de Quine). La conducta del informante nativo, incluida la conducta verbal, era la única clave para lo que estaba pasando dentro de su cabeza.

Muy pronto, en 1955, Rudolf Carnap (1956) tuvo una sugerencia para acabar con el problema. Su idea era emplear un ordenador (inteligente) como informante nativo. Para determinar lo que el ordenador sabe y cómo llega a saberlo, uno tiene acceso no sólo a su comportamiento (salidas) sino también, para cualquier detalle que se quiera, a su funcionamiento interno y a los sucesivos estados de su memoria. Uno puede usar esos datos para construir una teoría completamente empírica de *cómo llega a conocer*, y *en qué consiste su conocimiento* – una epistemología para ordenadores. Hoy, podemos llevar a cabo realmente esta empresa, y en un reciente artículo (Simon, 1992) yo esbocé cómo el concepto de *analiticidad* podía ser explicado por estos medios.

Pero puedo incluso proporcionar pruebas más concretas para la viabilidad de este tipo de exploración en epistemología. Los progresos humanos desde un estado de inocencia a un estado en el cual tienen algún dominio de un lenguaje natural y la habilidad para usar el lenguaje, para comprender y comunicar denotaciones referidas al mundo exterior a la cabeza del usuario. La tarea de *adquisición del lenguaje* les ha parecido a algunos tan difícil que han postulado una «capacidad para el lenguaje» innata para dar cuenta de tal adquisición. Esto evita, al menos en parte, la necesidad de resolver el problema epistemológico – en esta postura, estaba resuelto antes de nacer.

Sin embargo, si nosotros pudiésemos construir un programa de ordenador que, empezando como un recién nacido, realmente usara sus ojos, oídos y cerebro para adquirir un lenguaje natural, nos informaría sobre las condiciones previas para tal obtención – en qué consistiría una «capacidad para el lenguaje». Si el programa se iguala al aprendizaje inicial del lenguaje en humanos, entonces proporcionará una respuesta a esa parte, al menos, de la cuestión epistemológica de cómo conocemos.

A finales de los sesenta, Laurent Siklossy construyó dicho programa (Siklossy, 1972), el cual es por consiguiente al menos una primera aproximación a una teoría empírica del aprendizaje lingüístico humano. El programa se llama ZBIE (que no es un acrónimo de nada). ZBIE puede manifiestamente aprender al menos las partes más simples del vocabulario, la sintaxis y la semántica de cualquier lenguaje al que se le exponga y ha sido probado con inglés, alemán, francés y ruso.

Siklossy asume que el niño nace con, o adquiere con la edad cuando comienza el aprendizaje lingüístico, la habilidad de construir estructuras de símbolos en la memoria («imágenes mentales») de situaciones simples que aparecen ante sus ojos: el perro persiguiendo al gato, por ejemplo. Esta situación estaría contenida en la memoria como una estructura de símbolos consistente en dos patrones correspondiente a los objetos (el gato y el perro) incrustada en un patrón mayor correspondiente a la relación (persiguiendo) que los conecta. Estas estructuras son símbolos, cuya *denotación* es la escena externa. Esas estructuras son creadas en la memoria por procesos sensoriales y perceptivos que no tienen contenido lingüístico, sino que son consideradas mejor como «pictóricas». Corresponden a la habilidad que tienen los niños, antes de que comiencen a adquirir lenguaje, de reconocer objetos y situaciones corrientes cuando las ven.

Ahora bien, las situaciones codificadas son presentadas al sistema de Siklossy formando pareja con oraciones en el lenguaje que debe aprender. *Las oraciones están destinadas para denotar las correspondientes situaciones*. De este modo, ZBIE recibiría, con la estructura de símbolos denotando que el perro está persiguiendo al gato, la oración: EL PERRO PERSIGUE AL GATO. Después de que una serie de tales estímulos emparejados ha sido presentada, ZBIE habrá almacenado en la memoria (a) una red capaz de reconocer los objetos diversos que se ha encontrado y las situaciones diversas (relaciones) en las cuales se encuentran, (b) una red similar capaz de reconocer las palabras, en el idioma de que se trate, que se ha encontrado, (c) estructuras que recogen las reglas para organizar las palabras de varios tipos dentro de las oraciones gramaticales, y (d) enlaces que emparejan las palabras con los objetos y las relaciones que aquéllas denotan.

Supongamos, por ejemplo, que ZBIE ha asimilado PERRO, GATO, NIÑO, NIÑA, ACARICIA, PERSIGUE, VE, etc., y también tales escenas como el perro persiguiendo al gato, la niña persiguiendo al niño, y el niño acariciando al perro, junto con las frases que denotan esas escenas. Ahora una nueva escena, nunca antes encontrada por ZBIE, es presentada: la niña acariciando al gato. Al pedirlo, ZBIE producirá la frase, LA

NIÑA ACARICIA AL GATO. De este modo ZBIE satisface el primer requisito de Chomsky para un aprendiz o usuario de lenguajes naturales: ser capaz de entender y generar oraciones que previamente nunca se ha encontrado.

Para una información más completa de las capacidades y límites de ZBIE vea el informe de Siklossy (1972). Nuestro actual interés está en la luz que ZBIE arroja sobre los procesos simbólicos que están involucrados en el aprendizaje de un lenguaje. En concreto, incluso este simple ejemplo sirve como una refutación al argumento de la «Habitación China» de Searle (1984), de que los ordenadores no pueden entender el lenguaje. La «imposibilidad» en principio debe ceder ante la realización de hecho. A diferencia del sistema de la habitación china de Searle, ZBIE *posee enlaces entre sus palabras y las cosas del mundo-exterior y las relaciones que denotan.*

ZBIE deja también algunas cuestiones epistemológicas importantes sin contestar. Indica qué tipo de información semántica (imágenes mentales) sería necesario para respaldar el aprendizaje del lenguaje, pero no indica qué procesos sensoriales y perceptivos extraerían esta información a partir del mundo exterior. Esto es, por supuesto, una cuestión de gran interés en la investigación actual en inteligencia artificial en general y en robótica en particular. Mientras no ha habido una respuesta exhaustiva, existe hoy un número de sistemas robóticos que construyen estructuras de símbolos internos que representan las situaciones externas: por ejemplo, el sistema NAVLAB que usa sus propias capacidades sensoriales, interpretativas y motoras para conducir un vehículo que acelera hasta unas 50 millas por hora en una carretera.

Así pues tenemos la clase de respuesta en principio que necesitamos para los propósitos de la epistemología. Nosotros sólo necesitamos mirar a NAVLAB para ver cómo las útiles y veraces imágenes mentales de las situaciones externas pueden ser adquiridas y utilizadas. Esto no implica que la forma de trabajar de NAVLAB se parezca mucho a la forma humana en algún sentido; lo que NAVLAB demuestra es la existencia de mecanismos que pueden construir representaciones internas del mundo exterior que son útiles para guiar la acción.

El problema mente-cuerpo.

Con un gran número de programas existentes capaces de muchos tipos de tareas que, en humanos, llamamos pensamiento, y con pruebas pormenorizadas de que los procesos que algunos de esos programas usan reflejan estrechamente los procesos humanos observados, tenemos en la

mano una respuesta clara al problema mente-cuerpo: ¿cómo puede pensar la materia y cómo se relacionan los cerebros con los pensamientos?

La hipótesis del SSF afirma que los requisitos indispensables para pensar son patrones que pueden ser almacenados y manipulados. El conocimiento reside en el diseño de la materia, en combinación con los procesos que pueden crear y operar sobre tales patrones. *Un pensamiento sobre un gato,* quizá inducido por mirarlo o recordarlo, es una *estructura de símbolos* en esa parte del cerebro (probablemente el lóbulo frontal) donde reside «el ojo de la mente». Los procesos que operan sobre las estructuras de símbolos en este lugar pueden extraer, por ejemplo, información sobre subestructuras que podrían revelar el color del pelaje del gato, la longitud de sus bigotes, o la presencia o ausencia de un rabo. No hay nada misterioso en todo esto, pues los ordenadores pueden y llevan a cabo tales procesos. Si sus capacidades actuales para hacer tal cosa distan mucho de las capacidades humanas, la discrepancia entre ambos no implica ninguna laguna en nuestro conocimiento de los principios fundamentales involucrados.

Generalmente nosotros demostramos nuestro dominio de las leyes de la física realizando muy precisos pero simples experimentos de laboratorio en los cuales permitimos que sólo unas pocas cosas varíen y bloqueamos el resto del mundo lo mejor que podemos. El conocimiento de las leyes básicas que nosotros conseguimos a través de esta estrategia no garantiza la predecibilidad, mucho menos la constructibilidad, de fenómenos más complejos. Los meteorólogos pueden ser (y frecuentemente son) incapaces de predecir el tiempo, e incluso más frecuentemente son incapaces de decir algo sobre él. Esto no reduce nuestra confianza en que la atmósfera se comporta en conformidad con las leyes de la física. Mucha discusión confusa sobre la inteligencia artificial podría ser eliminada si le aplicáramos las mismas reglas para evaluar nuestro conocimiento como usamos en ciencias más antiguas.

Los límites de la computación

Nuestra discusión de la epistemología y del problema mente-cuerpo no ha tocado un asunto importante que frecuentemente surge en los debates sobre la validez de la hipótesis del sistema de símbolos físico. Incluso si fuese concedido que ciertas actividades humanas normalmente consideradas como suponiendo pensamiento pueden ser *simuladas* por un ordenador, quizá hay otras; cualitativamente diferentes, que no pueden serlo. La simulación cognitiva es a veces reclamada para trabajar sólo con «problemas de juegos» (por ejemplo, rompecabezas), o

«problemas de laboratorio» alejados del mundo diario, o problemas «bien estructurados» que no captan la vaguedad de las situaciones que los profesionales deben manejar. Ciertamente nuestros ejemplos descritos del aprendizaje del lenguaje y la percepción robótica van más allá de estos límites, pero no importa: ellos no garantizan que hayamos cubierto el territorio completo del pensamiento humano. Puede haber formas de pensamiento cualitativamente diferentes de aquéllos que han sido simulados.

Hay al menos dos líneas diferentes de argumentación que pretenden mostrar que la simulación por ordenador no puede recoger el pensamiento humano en todas sus formas. La primera objeción es que el ordenador es demasiado mecánico y «racional». El ordenador no puede tener errores («creativos» o de otro tipo) como la gente tiene y no puede pensar por caminos «no lineales» que los humanos pueden seguir, y los cuales a veces los llevan a las mejores ideas.

La segunda objeción (en realidad se trata de tres objeciones relacionadas) es que los ordenadores son incapaces de pensamiento «intuitivo», «perspicaz» o «creativo». Los dos grupos de objeciones no están totalmente desconectados, pero como plantean cuestiones un tanto diferentes los vamos a ver por turnos.

¿Es la computación mecánica y lógica?

Para cualquier definición aceptable de «mecanismo», el ordenador es un mecanismo. Pero la mayoría de los biólogos y probablemente la mayoría de los psicólogos cognitivos estarían de acuerdo en que el cerebro también lo es. Si por un mecanismo nosotros queremos decir un *sistema cuyo comportamiento en un punto en un momento está determinado por su estado interno actual combinado con las influencias que simultáneamente inciden sobre él desde fuera*, entonces cualquier sistema que pueda ser estudiado por los métodos de la ciencia es un mecanismo.

Pero el término «mecanismo» es usado también en un sentido más estrecho para referirse a *sistemas que tienen el relativamente fijo, rutinario, repetitivo comportamiento de la mayoría de las máquinas que vemos a nuestro alrededor.* Cualquier signo de espontaneidad exhibido por nuestro tostador, nuestra lavadora, nuestro automóvil, o la maquinaria de la fábrica va a ser atribuido a nuestras acciones sobre ellos (o, más recientemente, a las acciones de un ordenador que los controla).

Claramente el ordenador ocupa una posición ambigua aquí. Su conducta es más compleja, por orden de magnitud, que cualquier máquina que hayamos conocido, y no infrecuentemente nos sorprende, incluso

cuando está ejecutando un programa que nosotros hayamos escrito. Con todo, como reza el dicho, «sólo hace lo que tú le programas para hacer». Pero aunque lo que dice el tópico parezca ser así, es engañoso en dos aspectos. Es engañoso primero, porque a menudo se interpreta que significa: «sólo hace lo que tú crees que programaste que hiciera», lo cual claramente no es el caso.

Más grave, es engañoso porque da por sentado que los ordenadores y las personas son diferentes. Son diferentes (en esta dimensión) sólo si las personas se comportan de modo distinto a la manera como están programadas para comportarse. Pero si incluimos en el «programa» el estado completo de la memoria humana, entonces afirmar que las personas «no hacen sólo lo que están programadas para hacer» es equivalente a afirmar que los cerebros de las personas no son mecanismos, por lo tanto no son explicables por los métodos de la ciencia. No es una simulación por ordenador lo que está en juego, sino la posibilidad de explicar la conducta por completo.

Es mejor que dejemos a un lado los términos traicioneros «máquina» y «mecánico» y preguntemos más directamente en qué sentido un sistema programado puede exhibir espontaneidad. Por «espontaneidad» denotamos una conducta que es impredecible, quizá incluso para el propio sistema. Porque nosotros tenemos una capacidad muy limitada para predecir, por más que tomemos los intervalos más cortos, nuestras respuestas a nuestros pensamientos, mucho menos si tomamos nuestras respuestas a procesos que pasan subconscientemente en nuestras mentes, no es sorprendente que veamos nuestro comportamiento como que tiene grandes dosis de espontaneidad. Ni es sorprendente, considerando el conocimiento limitado que otros tienen de nuestros estados mentales, que nuestra conducta les aparezca incluso más espontánea a ellos.

Pero podemos decir lo mismo de los ordenadores. Si el ordenador es controlado por un programa de una gran complejidad en una tarea de resolución de problemas, nosotros probablemente no somos capaces de predecir mucho sobre qué es lo próximo que va a hacer. Muchos de nosotros hemos tenido la experiencia de jugar contra un programa de ajedrez y encararlo en términos cada vez más antropomórficos como que nos sorprende y nos amenaza por sus respuestas a nuestros movimientos. Los ordenadores sólo se comportan «mecánicamente» cuando ejecutan tareas monótonas de manejo de números, como invertir matrices o resolver ecuaciones diferenciales parciales. Entonces podemos ser capaces de detectar los ciclos repetitivos en su trabajo. En tareas de otros tipos que son problemáticas para las personas, su comportamiento es mucho menos simplemente diseñado.

Debemos también tomar en consideración la espontaneidad a corto y a largo plazo. A la larga, las personas pueden aprender. Ellas pueden cambiar sus programas. *Pero por supuesto los ordenadores también pueden aprender.* El programa EPAM discrimina entre objetos clasificándolos a través de una red, examinándolos en cada paso para enviarlos a diferentes ramas. EPAM tiene otros procesos que le permiten expandir la red cuando descubre que ha clasificado algo incorrectamente (dos estímulos diferentes, digamos, en el mismo nodo). Con la red expandida, ahora discrimina de forma diferente (y presumiblemente con más precisión) a como lo hacía antes. Ha aprendido.

Muchos mecanismos de aprendizaje, algunos de ellos tomando como modelo los procesos de aprendizaje humano como es EPAM, han sido incorporados en programas de ordenador. Entre los más interesantes desde un punto de vista psicológico están los *chunking* (combinación de trozos de información o procesos en unidades familiares más amplias), y los de *aprendizaje por ejemplos* (modificación de programas mediante el examen de los pasos por los cuales los problemas son resueltos y añadiendo nuevos procesos para hacer coincidir los pasos detectados en los ejemplos). Los mecanismos «chunking» proporcionan explicaciones para la automatización gradual de las habilidades humanas sumamente expertas, mientras que el aprendizaje por ejemplos se ajusta a procesos que han sido frecuentemente observados en situaciones escolares.

Nosotros concluimos que los ordenadores son capaces de los mismos tipos de espontaneidad que las personas son capaces, y que la cantidad de espontaneidad que ellos exhiban depende de la tarea en la cual están ocupados (como ocurre en las personas) y la complejidad del programa con el cual dirigen la tarea. En la medida en que nosotros tenemos éxito en la construcción de nuestros programas de simulación de la conducta de las personas, los ordenadores exhiben los mismos tipos y grados de espontaneidad que la gente tiene en las mismas tareas.

Comentarios similares pueden hacerse sobre la *racionalidad* de los ordenadores comparada con la de la gente. Con cuidado y suerte, nosotros podemos programar un ordenador para sumar una columna de cifras o multiplicar números sin error. *Si nosotros queremos simular la conducta humana, nosotros no haremos esto.* En vez de eso (como se ha hecho en el estudio de los «fallos» aritméticos de niños), *escribiremos programas que reproduzcan los tipos de errores que nuestros sujetos humanos cometen*, y de ese modo proporcionamos perspicacia en los fallos (desde un punto de vista lógico) de los programas humanos.

Déjenme proporcionar un ejemplo más sutil. Recientemente EPAM ha sido modificado para simular el comportamiento de una persona que

había aprendido (¡durante un período de tres años de práctica diaria!) a recordar secuencias de 100 números que se le habían leído a razón de un dígito por segundo. El sujeto humano sólo obtuvo éxito al recordar alrededor de la mitad de las listas; en las otras cometió uno o dos errores. Para simular al experto humano tuvimos que construir y comprobar una teoría del origen de los errores humanos y ajustarlo en el funcionamiento de EPAM. Además tuvimos que incorporar la fuente de los errores en el programa de EPAM de tal manera que EPAM continuara ajustándose a la conducta humana en los otros numerosos entornos de tareas en las cuales había sido previamente comprobado (Richman, Staszewski, y Simon, 1995). La meta no era mejorar el funcionamiento de EPAM sino entender y simular los límites del funcionamiento humano.

Viendo la cuestión de la racionalidad más generalmente, *un sistema es racional en la medida en que su comportamiento está bien adaptado para alcanzar sus metas sin exceso de tiempo o esfuerzo.* La economía ha intentado explicar la conducta humana en los mercados, con algún grado de éxito, asumiendo un imposible alto grado de racionalidad humana (*maximización de la utilidad esperada*). La teoría funciona, cuando lo hace, porque en situaciones que no son demasiado complejas la conducta de un sistema adaptativo estará moldeada exactamente por sus metas y la forma del entorno en el que está.

John Anderson (1990) se ha comprometido a mostrar que el mismo proceso se aplica a algunos fenómenos psicológicos. En este caso, el mecanismo de aprendizaje es evolución, no pensamiento. Durante largos períodos de tiempo, nosotros esperaríamos que los mecanismos evolutivos moldeen tanto las conductas como las estructuras de los organismos de tal manera que se adaptarán a sus entornos a la luz de sus metas y necesidades.

A medida que la complejidad aumenta, las desviaciones de la racionalidad perfecta llegan a hacerse más y más patentes, puesto que el organismo llega a ser menos capaz de calcular las respuestas óptimas. En entornos muy simples, podemos hacer buenas predicciones del comportamiento de los organismos adaptativos simplemente preguntando cuál sería la conducta más apropiada para alcanzar las metas de adaptación. En entornos más complejos, necesitamos tener en cuenta *las limitaciones computacionales del organismo* – los recursos, innatos o aprendidos, disponibles para descubrir los cursos de acción apropiados. La psicología cognitiva ha identificado los límites de la capacidad de la memoria a corto plazo, junto con los límites del conocimiento, como los dos parámetros internos más importantes que determinan cómo, y cómo de bien, una persona se adaptará a un entorno complejo.

La conducta humana pocas veces es perfectamente racional. Casi siempre es *limitadamente racional*, donde los límites de su destreza para encontrar los caminos óptimos son los límites de su conocimiento disponible y los límites de la habilidad humana para calcular las consecuencias de las acciones. En orden a programar ordenadores para simular el comportamiento humano frente a la complejidad cognitiva debemos incorporar las mismas limitaciones en la simulación por ordenador que hemos encontrado operativas en humanos. Si el programa es demasiado «lógico», si no es limitadamente racional, obviamente no pensará como lo hacen los sujetos humanos. Eso sería un fracaso en nuestra programación, no un límite intrínseco de la capacidad de los ordenadores para simular el pensamiento humano.

Producir programas de ordenador que sólo sean limitadamente racionales no es muy difícil. Si proponemos un problema que es muy extenso y al que le falta una estructura matemática simple y clara, generalmente no es posible producir un plan, incluso para el ordenador más poderoso, que encuentre una solución óptima. Si añadimos la condición de que la información disponible es radicalmente incompleta e inexacta, la limitación de la racionalidad está garantizada.

Deep Blue, el más poderoso programa de ajedrez hoy en día, es sólo limitadamente racional; no tiene ninguna forma de garantizar que ha encontrado los movimientos que son los mejores en el sentido de la teoría de juegos, aunque su éxito en competición con jugadores humanos muestra que generalmente encuentra movimientos muy buenos (movimientos mejores que la mayoría de los jugadores humanos pueden encontrar). Si nosotros ahora limitamos la velocidad y capacidad de computación de Deep Blue a un nivel más humano, su racionalidad llega a ser incluso más seriamente limitada. Para simular el juego de ajedrez humano, hemos intentado (con moderado éxito hasta la fecha) escribir programas que compensen el poder de computación que falta en los humanos con heurísticos selectivos que hagan la búsqueda más eficiente.

La misma observación se aplica a asuntos sobre el «pensamiento lineal» de los ordenadores. No está claro si la frase «pensamiento lineal» no tiene sino un significado metafórico, pero si es así o no, no hay nada sobre los ordenadores que requiera que piensen o más o menos «linealmente» que la gente. Todo depende del programa.

Intuición, Perspicacia, Creatividad.

El pensamiento humano, dice otro argumento, no es simplemente una cuestión de búsqueda, aunque sea selectiva, a través de un laberinto.

Hay formas más elegantes de pensamiento y éstas son requeridas en orden a descubrir y formular problemas, tener intuiciones sobre ellos y ganar perspicacia. Estas formas de pensamiento, bastante diferentes de aquéllas que han sido simuladas, se dice, aparecen cuando la gente está pensando creativamente. Una máquina, concluye el argumento, sólo puede hacer pensamiento mecánico, y por lo tanto, fallará cuando la intuición, perspicacia y creatividad sean requeridas.

Para investigar estas posibilidades, debemos tener definiciones de los términos claves: términos como «intuición», «perspicacia» y «creatividad». Si podemos proporcionar algún criterio para juzgar cuándo ha tenido lugar intuición, ha sido ganada perspicacia o el pensamiento ha sido creativo, entonces podemos investigar sobre los procesos que producen tales eventos.

Intuición.

Normalmente reconocemos la presencia de intuición cuando alguien resuelve un problema bastante rápidamente («instantáneamente») tras su presentación, y especialmente cuando él o ella no puede dar cuenta de cómo apareció la solución. («Justo de repente entró en mi mente»). Tú hablas con tu doctora, y después de describir unos pocos síntomas, ella proporciona el nombre latino de una enfermedad (y quizá algún consejo sobre el tratamiento). Le preguntas cómo lo sabe. «Es obvio, cualquier doctor/a competente lo reconocería inmediatamente».

La palabra «reconocer» es la clave. El proceso de resolución de problemas intuitivamente es indistinguible del proceso de reconocimiento de una persona u objeto familiar. El reconocimiento es repentino, y somos incapaces de describir exactamente qué características de la persona o del objeto llevaron a ello. A veces cometemos errores en el reconocimiento (léase: «nuestras intuiciones son falibles»). En realidad no es nuestro amigo sino un extraño, quien incluso no se parece mucho al amigo cuando se acerca. O la doctora respondió con la palabra latina equivocada: la enfermedad que sufrimos tiene un nombre diferente, pero las dos enfermedades tienen algunos síntomas en común.

Nuestros actos de reconocimiento o intuición llegan a ser más fiables cuando nosotros llegamos a estar más informados, a ser más *expertos*. Las pruebas que se han acumulado durante los últimos veinte años demuestran que la posesión de una gran, con numerosas entradas, «enciclopedia» de información sobre un dominio es la clave para la experiencia en ese dominio y para tener intuiciones fiables (reconocimientos) sobre los problemas en el dominio. (Para algunos indicadores de la

literatura exhaustiva sobre este asunto ver Charness 1989; y Ericsson & Staszewski, 1989). El índice para la enciclopedia consta precisamente de las entradas que serán reconocidas cuando los problemas se presenten al experto. El reconocimiento da acceso a la información asociada con la entrada en la memoria.

Los procesos de reconocimiento están modelados en detalle en el sistema EPAM, el cual proporciona una teoría, respaldada por pruebas empíricas extensas, de la conducta experta (Richman, Staszewski, Simon, 1995). Los procesos de reconocimiento también juegan un papel central en la mayoría de los sistemas expertos, algunos de los cuales imitan al menos en parte los procesos de los expertos humanos sobre los que ellos estaban moldeados.

Nosotros concluimos que ninguna forma de pensamiento más allá de aquellos sistemas de reconocimiento ya modelados como el EPAM son necesarios para explicar la intuición humana, su fiabilidad general en dominios de pericia, y su no fiabilidad frecuente fuera de esos dominios.

Perspicacia.

El término «perspicacia» a veces es usado casi sinónimamente con «intuición». Pero también lo usamos para referirnos a nuestra profundidad de comprensión de una situación, y especialmente para las maneras de representar la situación que produce la comprensión más profunda. Así Einstein consiguió una nueva perspicacia del problema del movimiento relativo cuando se dio cuenta (¿reconoció?) que para sincronizar relojes en diferentes lugares, las señales tienen que ser enviadas desde un lugar a otro. Esta perspicacia, junto con su creencia de que la velocidad de la luz sería la misma en cada marco de referencia, le llevó a la transformación de Lorentz y a la teoría de la relatividad especial.

En un estudio reciente, nosotros ayudábamos a los estudiantes a comprender la relatividad especial presentando las mismas descripciones que Einstein presentó en su escrito original de 1905 (Qin & Simon, 1992). Dibujando un diagrama simple, ellos eran capaces de calcular el tiempo que tardaría un rayo de luz para ir desde el extremo de una varilla al otro y volver, para condiciones diferentes del movimiento de la varilla. Combinando las ecuaciones para una varilla en movimiento y una varilla inmóvil, ellos eran capaces de establecer la ecuación desde la cual son derivadas las ecuaciones de Lorentz. Las descripciones de Einstein (¡él no publicó diagramas en su artículo!) les condujeron a la perspicacia que les permitió entender la relación entre el tiempo y el espacio coordinados en los dos marcos de referencia.

Consideremos un ejemplo menos esotérico: el tablero de damas mutilado, un problema de IA propuesto por primera vez, creo, por John McCarthy (Kaplan & Simon, 1990). Tenemos un tablero de damas, y 32 fichas de dominó, cada una capaz de cubrir exactamente dos cuadros adyacentes del tablero. Obviamente, podemos cubrir el tablero completamente con las fichas del dominó. Ahora bien, los cuadros noroeste y sudeste del tablero son suprimidos ¿Podemos cubrir los 62 cuadros restantes con 31 fichas de dominó?

A los sujetos de laboratorio se les da este problema para trabajar con él, cada vez con mayor frustración, por un largo periodo de tiempo. Ellos intentan varios cubrimientos, siempre sin éxito. A veces trabajan sobre un tablero 4 x 4 simplificado (también sin éxito). Ocasionalmente, y casi siempre después de varias horas de fracasos, un sujeto notará que los cuadros dejados sin cubrir después de un intento sin éxito al cubrir el tablero son siempre del mismo color – el color opuesto al color de los dos cuadros que fueron quitados. Muy frecuentemente, los sujetos que notan este hecho resuelven el problema en dos o tres minutos. En sus protocolos verbales ellos dicen: «¡Oh! El tablero mutilado tiene menos cuadros rojos que negros. Pero cada ficha de dominó cubre un cuadro rojo y uno negro, así que será imposible cubrir más de un color que de otro. El problema no tiene solución».

La perspicacia aquí consiste en ver que sólo el número de cuadros de cada color es importante, y no el orden de las fichas de dominó sobre el tablero. Si no hay el mismo número de cada color, entonces no hay manera de cubrirlos con las fichas de dominó, las cuales siempre cubren un cuadro de cada color. Otra vez, esta perspicacia aparece asociada muy de cerca con un acto de reconocimiento: el reconocimiento de que cada intento de cubrir el tablero deja dos cuadros del mismo color sin cubrir.

Esta interpretación puede ser apoyada repitiendo el experimento con algunos cambios: suministrando un tablero con los cuadros sin colorear; suministrando un tablero en el cual los cuadros alternos son etiquetados «pan» y «mantequilla». La primera manipulación suprime una clave que puede llevar a reconocer la falta de igualdad de colores en el tablero mutilado. La segunda manipulación llama la atención de la paridad de cuadros contiguos mediante etiquetas que no tienen otro significado. Como se había previsto, en el experimento la primera manipulación reduce la probabilidad de resolver el problema en un tiempo dado; la segunda la incrementa.

Kaplan (comunicación personal) ha escrito un programa de ordenador que busca las propiedades que permanecen invariables cuando se intentan las soluciones del problema. Agrega tales propiedades a las re-

presentaciones de los objetos en el problema (añade el color a la descripción de las fichas de dominó y los cuadrados). Previamente, podía sólo comparar el número total de cuadros a cubrir con el número de fichas de dominó. Ahora compara el número de cada color cubierto por las fichas de dominó con el número de cada color en el tablero, y de ese modo descubre la imposibilidad.

Una fuente importante de perspicacia en el descubrimiento científico es la sorpresa. Se pueden citar muchos ejemplos de descubrimientos de enorme importancia que comenzaron con una sorpresa (por ejemplo: Roentgen, rayos X; los Curies, radio; Tswett, cromatografía; Fleming, penicilina; Krebs, ciclo de la urea; Faraday, la inducción electromagnética, y muchos otros). En cada caso, la sorpresa condujo a una perspicacia importante, pero ¿qué son los mecanismos de sorpresa y perspicacia?

Kulkarni construyó un programa, llamado KEKADA, que planifica estrategias experimentales para enfrentarse con los problemas científicos (Kulkarni & Simon 1988). Comenzando con dos enunciados (más o menos precisos) de la meta y los métodos de experimentación disponibles, propone un experimento. Sobre la base del resultado del experimento, propone otro, y así sucesivamente. Crea expectativas, sobre la base del conocimiento y la experiencia previos, sobre los resultados de los experimentos que propone. Si estas expectativas son incumplidas, experimenta «sorpresa», y comienza a planificar nuevos experimentos para definir el ámbito del fenómeno sorprendente y los mecanismos que podrían producirlo. Usando esta estrategia, tuvo éxito en la simulación del programa experimental de Krebs para el descubrimiento del camino de la síntesis de la urea, en el de Faraday para explicar la producción de corriente eléctrica por un imán en movimiento, y varios otros.

Los mecanismos que subyacen al funcionamiento de KEKADA son familiares. En la medida en que posee conocimiento sobre un dominio, KEKADA puede formar expectativas sobre el resultado de los experimentos. Teniendo expectativas formadas, se puede sorprender si las expectativas son frustradas. («Los accidentes le ocurren a la mente preparada» – Pasteur). Usando su conocimiento experto, puede planificar experimentos para explicar la sorpresa. Así que las perspicacias obtenidas investigando una sorpresa dependen otra vez de las poderosas capacidades para reconocer claves familiares.

Los programas como KEKADA enseñan una importante lección metodológica sobre la investigación de procesos cognitivos. Ya que no podemos poner a Krebs o Faraday en un laboratorio, ni siquiera entrevistarlos en la actualidad, ¿cómo podemos comprobar si los procesos que el programa de ordenador está utilizando se parecen a los procesos

que ellos usaban para hacer sus descubrimientos? Ciertamente no hay modo alguno en el que podamos comprobar tales simulaciones con la resolución temporal que está disponible cuando podemos tomar protocolos verbales.

Sin embargo, en los casos de Krebs y Faraday, tenemos disponibles sus cuadernos de laboratorio, los cuales proporcionan una lista completa de los experimentos que realizaron – normalmente varios cada día. Nosotros podemos comparar los experimentos que ellos llevaron a cabo con aquéllos propuestos por KEKADA, ajustándolos ambos con respecto al contenido y la secuencia temporal. Esto nos da un cuerpo sustancial de datos (aunque con intervalos de tiempo de días en vez de minutos) para evaluar la teoría. Podemos también examinar las publicaciones científicas para sus consideraciones (retrospectivas) de los razonamientos que emplearon; y en el caso de Krebs, también tenemos entrevistas retrospectivas que el historiador de la ciencia, C.L. Holmes, realizó antes de la muerte de Krebs.

Cuando los datos de esa clase están disponibles, podemos comprobar nuestros modelos comparándolos con los descubrimientos de gran interés histórico e importancia. Podemos tener alguna seguridad de que si el más poderoso e imaginativo pensamiento requiere intuición y perspicacia, entonces esas cualidades deben estar presentes en los eventos psicológicos que guían tales descubrimientos, y presentes también en los programas de ordenador que simulan esos eventos, si bien sólo en una escala de días más bien que en minutos.

Creatividad.

La creatividad quizá sea un término incluso menos preciso que intuición o perspicacia. Una acción o su producto se contempla como creativa en la medida en que el producto tiene valor a lo largo de alguna dimensión (estética, científica, económica, etc.) y en la medida en que es novedoso. La novedad valiosa es la marca de la creatividad. Nótese que el criterio se refiere al resultado, no a los procesos. Pero quizá hay procesos especiales que son conducentes a producir novedad valiosa.

En nuestra discusión de la teoría de la relatividad y las estrategias experimentales que KEKADA modela ya hemos entrado dentro de la esfera de la creatividad. Sería más bien algo excéntrico pretender que Einstein, Krebs y Faraday no eran creativos. Pero miremos un poco más allá para ver si otros aspectos de la creatividad pueden haber sido iluminados por los intentos de simular el descubrimiento científico.

Formular problemas y proveerlos con representaciones eficaces a menudo es mencionado como una clase importante de creatividad. Ya hemos tenido al menos un vislumbre, en el ejemplo del tablero de damas mutilado, de cómo nuevas representaciones podrían ser descubiertas para hacer un problema difícil soluble. Relacionados muy de cerca con las nuevas representaciones están los *nuevos conceptos*. Conceptos como momento, energía y masa no estaban simplemente «ahí» en la naturaleza. Ellos surgen a base de un gran esfuerzo humano (en este caso, el esfuerzo de figuras tales como Descartes, Huygens, y Newton) en respuesta a problemas de caracterización de los fenómenos reales. ¿Qué sería necesario para simular el descubrimiento de nuevos conceptos de estos tipos?

No tenemos que especular sobre la respuesta a esta pregunta, porque el programa BACON, entre otros, ya tiene la capacidad de construir conceptos teóricos nuevos (conceptos para denotar cosas que no son directamente observables) en el curso de construcción de teorías para describir los datos (Langley et al., 1987). BACON es un sistema de descubrimiento de leyes dirigido por datos. Dados algunos datos desde la experimentación o la observación, trata de encontrar una ley algebraica que describirá los datos parcamente. Dados los datos sobre las masas y temperaturas de dos frasquitos de agua y la temperatura de equilibrio cuando son mezclados, llega a la ley de Black, la cual dice que la temperatura de equilibrio es la media de las temperaturas de los dos líquidos componentes, ponderados por sus respectivas masas.

Pero ¿qué dice si los líquidos mezclados son diferentes, agua y alcohol? Con un poco más de dificultad, BACON descubrirá que la temperatura de equilibrio es todavía una media ponderada de las temperaturas de los componentes, pero los pesos no son ahora simplemente las masas, sino las masas multiplicadas por una característica constante para cada líquido (digamos, w para el agua y a para el alcohol). Esas constantes, primero descubiertas por Joseph Black y posterior e independientemente por BACON, son conocidas como los *calores específicos* de las respectivas sustancias.

Desde este ejemplo, vemos que el enriquecimiento de la representación de un problema por la introducción de nuevos conceptos puede también ser simulado. Nosotros no tenemos ninguna prueba muy buena, hasta el momento, de que los procesos que BACON utiliza para llevar a cabo esto estén cercanos a los procesos utilizados por los descubridores humanos, pero los procesos de BACON se construyen sobre heurísticos más bien simples bastante similares a los heurísticos que hemos observado que los humanos usan en otras situaciones.

¿Es diferente el lenguaje natural?

Casi todos los ejemplos que he proporcionado de la simulación de la intuición, perspicacia y creatividad han sido sacados de dominios científicos. Las personas que no son científicas a veces son reacias a creer que esos procesos ocurren en tales dominios, porque la ciencia se supone que es ordenada y «lógica». Aquellos de nosotros quienes gastamos nuestras vidas en la ciencia lo sabemos mejor, pero quizá sería útil despejar las dudas girando a un dominio bastante diferente en el cual la gente puede exhibir su intuición, perspicacia y creatividad. Déjenme preguntar cómo podríamos abordar la simulación del pensamiento humano al leer un texto.

La maquinaria que nosotros necesitaríamos para esta tarea obviamente incluiría un léxico con entradas (por ejemplo, una memoria como la de EPAM) y un analizador sintáctico. Los analizadores sintácticos sabemos cómo se construyen, aunque quizá no al nivel de sofisticación que algunos humanos alcanzan. De hecho, ZBIE ya nos ha mostrado cómo el aprendizaje del lenguaje humano (simple) puede ser simulado. El problema de mayor dificultad parecería ser cómo extraer significados desde el texto que está hecho de palabras en nuestro léxico y que podemos analizar sintácticamente.

Cómo uno extrae significados de los textos, o incluso si los textos *tienen* significados, es un asunto vivo en el presente en el campo de la crítica literaria. Jerry Hobbs ha proporcionado una respuesta muy convincente a la pregunta qué son los significados. Él dice (Hobbs, 1990) que el significado de un texto no es una función de una sola variable, el texto, sino de dos variables, el texto y la mente del lector, $S = f(T, L)$. Si sustituimos un texto por otro, obtenemos un significado diferente para el mismo lector; y si sustituimos un lector por otro, obtenemos un significado diferente para el mismo texto. Creo que hemos sabido eso desde hace mucho tiempo. Los textos llegan a estar muy cargados de cultura, no sólo porque el autor está inmerso en una cultura, sino también porque los lectores están inmersos en la misma o en otras culturas.

Podemos dibujar una imagen aproximada sobre qué sucede en los términos de los mecanismos que ya hemos considerado, en particular, los mecanismos de reconocimiento y los mecanismos de inferencia. Déjenme ilustrar esto brevemente con el pasaje de apertura de una de las novelas cortas de Camus, *La Chute:*

> ¿Puedo yo Señor, ofrecerle mis servicios sin riesgo de ser importuno? Me temo que usted puede no saber cómo hacer que el estimable gorila que

> preside los destinos de este establecimiento le entienda. De hecho, el habla sólo holandés. A menos que me permita interceder por usted, él no acertará que usted quiere ginebra.

¿Qué podría evocar el reconocimiento de este texto? Por la frase tercera llegamos a saber que estamos en Holanda, quizá en Amsterdam. Por supuesto «Amsterdam» puede evocar ahora toda la masa de información que tengamos sobre esa ciudad, y las asociaciones con ella. «Ginebra», a su vez, hace remisión a «bar», determinando el entorno e identificando al «gorila» como un camarero. Otras conexiones posibles son más sutiles. «Interceder por usted» sugiere la ley, y de hecho anuncia la profesión del hablante. Y la palabra «importuno», aplicada a la relación del hablante con el que escucha, puede conducir a algunos lectores al marinero antiguo. Además, el tono del pasaje puede evocar en el lector cualquier afecto que esté asociado en la memoria del lector con un comportamiento cortés (o ¿pretencioso?).

Vemos aquí unas series completas de reconocimiento evocando asociaciones ya almacenadas en la memoria. Aunque una idea sugiere otra, hay muy poco aquí que un lógico reconocería como un «razonamiento». Que si la ginebra es mencionada, ¿estamos en un bar? Sería más preciso decir que el bar es *sugerido* por la ginebra que decir que el bar es *inferido* desde ella.

¿Qué hay del autor? ¿Qué podemos decir de sus procesos mentales? Sin más datos que este corto pasaje, sólo algunas conjeturas. Es de suponer que él comenzó con una meta, quizá la de proporcionar un escenario y un estado de ánimo para la historia sobre la que va a hablar. Su memoria le proporciona a él una descripción concreta de un tipo particular de establecimiento (que él dibuja muy probablemente desde su propia experiencia). Él escoge de su almacén de información una serie de palabras y frases que evocarán en el lector la imagen de un bar, posiblemente en un vecindario de clase baja. Él está contando con que esas palabras evoquen en la mente del lector algo de las mismas asociaciones que están presentes en su propia mente. O bien él tiene algún sentido de lo que sus lectores conocen o bien asume que sus almacenes de conocimiento se parecen a los suyos.

No intentaré llevar el análisis más allá. Pero quizá ya se sugiera que los procesos que nosotros estamos observando aquí, de ambos, lector y escritor, no son distintos de los procesos que hemos visto en otros tipos de pensamiento, particularmente de aquellos procesos de reconocimiento que hemos asociado con «intuición» y «perspicacia». Al igual que

podemos «darle un truco» a un sujeto en el problema del tablero de damas mutilado presentándole una pista que atraiga la atención a los colores alternos de los cuadros del tablero, así Camus «da un truco» al lector asignando rasgos personales al hablante a través de su manera de describirlo o reproducir su discurso.

Un breve, esbozadamente analizado ejemplo no es una demostración. Pero espero que mi discusión del pasaje de *La Chute* al menos plantee la posibilidad de que la actividad mental, la actividad mental enteramente creativa, fuera de las ciencias puede producir el mismo análisis que los dominios científicos estudiados más meticulosamente que yo he dibujado antes en mis otros ejemplos.

Conclusión

En este artículo he explorado, algunas de las características centrales de una teoría computacional de la cognición, y especialmente el tipo de teoría que ha sido asociada con la hipótesis del sistema de símbolos físico. Yo he buscado ilustrar cómo un modelo computacional puede ser usado para dirigir cuestiones epistemológicas en filosofía y el problema mente-cuerpo.

Mi foco principal, sin embargo, ha sido sobre la tesis de que el pensamiento humano puede ser representado por tal teoría computacional. En particular, he comentado arriba algunas de las principales objeciones que han sido alzadas contra esta tesis, especialmente el asunto de que el computacionismo es «mecánico», y por consiguiente incapaz de representar tales tipos importantes de procesos del pensamiento humano como los no-numéricos, los intuitivos, los perspicaces, los creativos, los propensos al error y los no lógicos. Principalmente por señalar ejemplos de simulaciones por ordenador de procesos humanos que realmente han sido construidos y comparados con datos humanos, me he comprometido a mostrar que todas estas características humanas pueden ser, y han sido simuladas.

Es un truismo que la mente humana es compleja. Así como que es de naturaleza física y biológica. La complejidad no ha impedido que los científicos físicos y biológicos hayan identificado muchos de los mecanismos principales que están debajo de los procesos físicos y biológicos, incluso aunque ellos rara vez puedan mostrar con detalle justo cómo estos mecanismos operan en escenarios del mundo real complejos (por ejemplo: no pueden predecir exactamente o en detalle cómo la atmósfera produce el tiempo diario del mundo, o cómo las leyes de la mecánica gobiernan la estabilidad o inestabilidad de la tierra sometida a las sacudidas de un terremoto).

De la misma manera, la psicología del procesamiento de la información ha identificado muchos de los principales mecanismos que los seres humanos usan para abrirse camino en un mundo cuya complejidad posee una magnitud demasiado grande para que la racionalidad limitada del ser humano pueda manejarla con alguna exactitud. No puede predecir en ningún detalle cómo una persona particular manejará una decisión de negocios particular compleja, o cómo una crisis de misiles cubanos será resuelta. Esto no implica que los mecanismos que gobiernan tales eventos sean diferentes de los que gobiernan las situaciones más simples en el laboratorio. El funcionamiento de los mecanismos de los que yo he hablado aquí ha sido demostrado en una amplia variedad de tareas, la mayoría aunque no todos ellos relativamente sencillos y hay razones sobradas para suponer que operan de la misma manera en las tareas que todavía no han sido simuladas en detalle.

Referencias

Anderson, J. R. (1983). *The architecture of complexity.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Carnap, R. (1956). *Meaning and necessity.* Chicago, IL: University of Chicago Press.

Charness, N. (1989). Expertise in chess and bridge. En D. Klahr y K. Kotovsky, (Eds.) *Complex information processing.* Hillsdale, NJ: Erlbaum.

Ericsson, K. A., y Staszewski, J. J. (1989). Skilled memory and expertise. En D. Klahr y K. Kotovsky (Eds.), *Complex information processing.* Hillsdale, NJ: Erlbaum.

Hobbs, J. R. (1990) *Literature and cognition.* Stanford, CA: Center for the Study of Language and Information, Stanford University.

Kaplan, C. A., y Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22,* 374-419.

Kulkarni, D., y Simon, H. A. (1988) The processes of scientific discovery: The strategy of experimentation. *Cognitive Science, 12,* 139-176.

Langley, P., Simon, H. A., Bradshaw, G. L. y Zytkow, J. M. (1987). *Scientific Discovery.* Cambridge, MA: MIT Press.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard Univesity Press.

Newell, A., y Simon, H. A. (1972). *Humam problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Newell, A., y Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM, 19,* 113-126.

Qin, Y., y Simon, H. A. (1992). Imagery as process representation in problem solving. Proceedings of the 14th Annual Conference of the Cognitive Science Society, 29 Julio – 1 Agosto, 1992.

Quine, W. V. (1960). *Word and object.* Cambridge, MA: Harvard University Press.

Richman, H. B.: Staszewski, J. J., y Simon, H. A. (1995). Simulation of expert memory using EPAM IV. *Psychological Review, 102,* 305-330.

Rumelhart, D. E. y McClelland, J. L. (1986). *Parallel distributed processing.* Cambridge, MA: MIT Press.

Searle, J. R. (1984). *Minds, brains, and science.* Cambridge, MA: Harvard Univesity Press.

Siklossy, L. (1972). Natural Language learning by computers. En H. A. Simon y L. Siklossy (Eds.), *Representation and meaning.* Englewood Cliffs, NJ: Prentice-Hall.

Simon, H.A. (1979,1989). *Models of thought* (Vol. 1 y 2). New Haven, CT: Yale University Press.

Simon, H.A. (1992). The computer as a laboratory for epistemology. En L. Burkholder (Ed.), *Philosophy and the computer.* Boulder, CO: Westview Press.

Vera, A., y Simon, H. A. (1994). Reply to Touretsky and Pomerleau: Reconstructing physical symbol systems. *Cognitive Science, 18,* 355-360.