Claudio Gutiérrez

PERSPECTIVAS DE LAS MAQUINAS INTELIGENTES O LA PSICOLOGIA DE LOS COMPUTADORES

Summary: *Is Aristotle's definition of man as the rational animal still valid in an information era in which AI computer programs not only solve equations, but play chess, do medical diagnosis and discover mathematical concepts, or are we rather "emotional machines"? The physical-symbol-systems theory of intelligence offers a generalized explanation of intelligence and a new paradigm for Psychology. For the fist time in history philosophic studies of the mind can benefit from technical attempts to build a mind.*

Resumen: *¿Es la definición aristotélica del hombre como animal racional todavía válida cuando programas de IA resuelven ecuaciones, juegan ajedrez y hacen diagnóstico médico, o más bien somos "máquinas afectivas"? La teoría de los sistemas de símbolos físicos ofrece una explicación generalizada de la inteligencia y un nuevo paradigma para la Psicología. Por primera vez en la historia los estudios filosóficos de la mente pueden beneficiarse de los intentos técnicos de construir una mente.*

En la Antigüedad, en tiempo de Platón y Aristóteles, se consideraba que la actividad más noble a que podía dedicarse el hombre era la contemplación de las Matemáticas. Hoy día existen máquinas--computadores debidamente programados--que resuelven polinomios mejor que los hombres. ¿Deberemos considerarlas, a pesar de no ser de carne ni de sangre, seres filosóficos, dignos de ser admitidos a la Academia o al Liceo? Si a eso agregamos que desde hace 30 años se cultiva en el mundo una disciplina nueva llamada Inteligencia Artificial (IA), que trata de dotar a los computadores de la capacidad de razonar, no numéricamente sino de manera cualitativa, el problema se hace aun más serio. Porque entonces, y para seguir situados en la perspectiva helénica, ¿qué sucede con la definición del hombre como animal racional? ¿Se convierte solamente en una manera de emparentarnos con el computador, *artefacto racional*? A este respecto es revelador el trabajo de Sherry Turkle (TURKLE 84) con niños norteamericanos familiarizados con el uso de computadores; a la pregunta por la diferencia entre ellos y el computador una respuesta unánime no se deja esperar: los computadores no sienten, mientras que ellos sí. Todo pareciera indicar que la cultura contemporánea está lista para aceptar una nueva definición del hombre; no "animal racional" como propusiera Aristóteles, sino "maquina afectiva". De acuerdo con esto, en unos cuantos decenios, la superficie de la tierra estará compartida por seres racionales, unos de los cuales habrán nacido de mujer y otros habrán sido manufacturados, por hombres o por otros individuos de su mismo género. ¿Afectará esta nueva perspectiva la autoimagen del hombre?

El concepto de hombre ha pasado en la historia contemporánea por diversas devaluaciones. La primera gran devaluación sucede cuando a Copérnico y a Galileo se les ocurre disputar que la tierra sea el centro del universo. Un buen número de consecuencias importantes se siguen de este cambio en autoimagen: en religión, ciencia, arte y literatura;

dan lugar a todo un movimiento cultural que llamamos Renacimiento. La segunda devaluación ocurre en el siglo pasado, cuando Darwin enseña que la aparición del hombre no se debe a un acto de creación especial, sino ha sido producto de una evolución ciega regida por el principio de selección natural. También aquí se siguen muchas consecuencias en la filosofía y las relaciones sociales de la época. La tercera devaluación sucede como resultado de la obra combinada de Marx y Freud, los cuales explican la obra intelectual social e individual de los hombres como resultado de fuerzas no racionales. El descubrimiento de Freud de que la mayoría de las actividades de la mente son inconscientes, nos hace quedar mejor definidos como seres racionalizantes que como seres racionales: más que ejercer la razón, lo que hacemos es dejarnos llevar por nuestros impulsos inconscientes para construir retroactivamente justificaciones intelectuales de lo actuado. En cuanto a Marx, le toca demostrar que gran parte de las cosas en que una sociedad cree, en realidad no las cree por su mérito intelectual intrínseco, sino porque la clase dominante en esa sociedad se ve favorecida por esas creencias que más que ideas son ideología. Esta incluye contenidos intelectuales tan elevados como el arte, la literatura, la música, la filosofía o la religión. Aquella parte de la cultura que consideráramos más excelsa, la obra del espíritu, resulta ahora imposición de la clase dominante e instrumento de opresión. Otra vez aquí, el hombre es redefinido: como ser ideologizante más que como ser racional. El análisis de Marx logra lo mismo en el plano social que el análisis de Freud en el plano individual.

Ahora nos toca vivir la cuarta devaluación, por obra de la Informática. La devaluación informática consiste en reconocer que en vez de ser animales racionales como creía Aristóteles—racionalizantes (Freud); ideologizantes (Marx)—, somos más bien máquinas afectivas. Las máquinas y los hombres tenemos en común la razón, y lo que nos diferencia de las otras máquinas es que además tenemos afectos, como antes decíamos tener en común con los animales los afectos y diferenciarnos de ellos por la razón. Pero es mucho más discutible que el que tengamos afectos nos dé un privilegio tan grande sobre las máquinas, como el que suponíamos nos daba la razón sobre los animales, puesto que los afectos distorsionan nuestra conducta y suelen conducirnos a menudo a acciones que nos perjudican. Pareciera que tendremos que aceptar, cualquiera que sea nuestra posición sobre cuestiones metafísicas, que ese algo que llamamos inteligencia que a través de los milenios ha parecido ser patrimonio exclusivo de los hombres, ahora ya no lo será tanto. No tendremos ya el monopolio de la inteligencia, a menos que redefinamos la inteligencia de una manera tan caprichosa como para no incluir en ella la capacidad de resolver ecuaciones matemáticas, o de jugar ajedrez, o de hacer diagnóstico médico, o de descubrir y probar teoremas lógicos o geométricos.

El surgimiento del tema de la IA como cuestión científica seria, según Margaret A. Boden (BODEN 84), puede fecharse en 1950, justo cuando A. Turing escribe en *Mind* sobre la posibilidad de que las máquinas adquieran inteligencia. Aparte de rebatir diversos argumentos que se han dado para demostrar la imposibilidad de que las máquinas lleguen a ser inteligentes, ese artículo propone una prueba para decidir si una máquina ya ha alcanzado esa meta. La prueba (desde entonces conocida como "la prueba de Turing") consiste en adaptar un juego de salón a la interacción hombre/máquina; si la persona cree tener como interlocutor (a través de un teletipo) a otra persona, cuando en realidad tiene a una máquina que se comporta como lo haría una persona, entonces podremos decir que la máquina es inteligente. El juego consiste en adivinar el sexo de un jugador, invisible y empeñado en engañar a su adversario. Implica no solo la capacidad de entender estereotipos sexuales, sino también la habilidad de manipular los mecanismos de la simulación y el engaño. Es muy revelador que Turing prefiriera escoger esta prueba como criterio de inteligencia, en vez de utilizar algún recurso que midiera experticia, como habría sido por ejemplo la capacidad de jugar ajedrez. El tiempo le ha dado la razón: en 1987, varios programas de computación juegan ajedrez a nivel de maestro y están peligrosamente cerca de disputar el campeonato de USA; en cambio, no hay todavía ninguna máquina capaz ni lejanamente de engañar a un ser humano haciéndose pasar por mujer.

Todos estos acontecimientos dan base para pensar que la investigación en IA ha hecho surgir un nuevo paradigma para la Psicología, por lo menos para la rama de esta ciencia que se ocupa de los problemas del conocimiento, la Psicología Cognoscitiva. Tal paradigma corresponde a un esfuerzo para definir y analizar el fenómeno del conocimiento intelectual con total universalidad, sin tomar en cuenta en qué tipo de ente se manifiesta.

Cabe advertir que los investigadores en IA no son los primeros en enfocar el problema de la inteligencia con un grado de generalidad que trasciende la extensión del género humano. Los psicólogos conductistas nos ofrecen con anterioridad un paradigma de interpretación del comportamiento en que no se distingue entre sistemas inteligentes humanos y animales. El conductismo trata de entender los fenómenos psicológicos humanos partiendo de principios muy generales obtenidos en el estudio de la conducta de los animales, a partir de los experimentos de Favlov. En esto actúan haciendo aplicación de una famosa máxima de método científico que debemos a Francis Bacon: "Busca percibir en la simplicidad de los fenómenos más sencillos las leyes universales que actúan en los fenómenos más complejos". La misma situación vuelve a ocurrir con los modelos de IA, pero en este caso los fenómenos más sencillos se encuentran en el funcionamiento de las máquinas, no en el comportamiento de los animales.

Podemos decir que a esta altura del clima intelectual hay consenso entre los autores en que el paradigma computacional es definitivamente superior al paradigma conductista como marco filosófico para interpretar la conducta humana. ¿Por qué? ¿En qué consiste el contraste entre ambos? El conductismo es un heredero del empirismo inglés de Locke y Hume, pero específicamente se caracteriza por insistir en que la única manera de conocer el comportamiento es por sus manifestaciones externas. La introspección cae en el descrédito; se generaliza la prohibición de hablar de procesos mentales; solo lo externo, lo observable, es materia de estudio científico. En oposición a esto, el paradigma computacional reivindica los procesos internos, puesto que hay un estado interno en la máquina; además hay entrada y salida que corresponden al estímulo y respuesta del conductismo. Los conductistas son ciegos a la etapa intermedia entre la entrada y la salida porque no tienen ninguna analogía empírica a qué apelar para esa etapa. Los computacionistas sí: tenemos un proceso interno en la máquina, si no directamente observable en el momento de su funcionamiento, empíricamente asegurado por ser resultado del proceso mismo de su construcción. Si aceptamos la existencia de esos estados intermedios entre estímulo y respuesta, entrada y salida, la actitud del científico del comportamiento se transforma: reconoce que la base de lo mental y de la inteligencia son estados internos, que pueden ser interpretados como representaciones de objetos y acontecimientos del mundo exterior. Una vez comprendido esto, podemos hablar de inteligencia humana, de inteligencia animal, de inteligencia de máquina, o de inteligencia extraterrestre, como capacidad de manipular esas representaciones. Se abre así la posibilidad de elaborar una teoría de la inteligencia en general, independiente de su "encarnación" en humano, animal, máquina o marciano.

Entre los más creativos de los articuladores del paradigma computacional debemos destacar a A. Newell y H. Simon. Ante todo, deben ser mencionados por su labor pionera de los años cincuentas, con la creación de un programa muy influyente, el Solucionador General de Problemas ("General Problem Solver", mejor conocido como GPS); con él tratan de emular los métodos más generales de la inteligencia humana, lo que comúnmente identificamos como sentido común. Fracasaran en esa empresa, pero tal fracaso les dará ocasión para formular las primeras leyes del nuevo paradigma y, eventualmente, los llevará a formular su teoría fundamental. El GPS tiene tres objetivos: el primero es construir un modelo de la inteligencia humana; es un interés científico, de comprensión del fenómeno; el segundo es un interés tecnologico: crear herramientas inteligentes, donde lo importante es el provecho que se pueda obtener, el tercero, de fuste más bien filosófico, es descubrir en qué consiste la intèligencia en general, independiente de su incorporación en un ser humano o en cualquier otro organismo o mecanismo. Muchos años después, con ocasión de recibir una distinción cobijada por el nombre de Turing (hermosa coincidencia), Newell y Simon proponen la *teoría de los sistemas físicos de símbolos* como explicación última de la inteligencia. "Símbolos" es otra forma de llamar a las representaciones; "físicos" subraya su carácter material, y por ende su fundamento empírico; "sistema" alude a la inmensa complejidad de los fenómenos involucrados. La teoría insiste en que el uso del computador nos ha puesto en contacto, por primera vez, con la naturaleza física dea inteligencia. Antes del computador, la idea de símbolo no era ni siquiera inteligible sin una inteligencia que interpretara el símbolo: lo simbólico era simbólico *para alguien*. Ahora las cosas se invierten: el simbolismo (de naturaleza física) es la base para entender la inteligencia. Esta teoría es la más clara expresión del paradigma computacional para las ciencias del comportamiento.

Los lenguajes de computación—especialmente los más avanzados, como el LISP, el favorito de los investigadores en IA— están constituidos por dos clases de símbolos: los que representan procesos o acciones, y los que representan valores (es decir, otros símbolos). El concepto de símbolo adquiere en computación un significado muy concreto: es un puntero, una flecha almacenada en un lugar de la memoria que señala otro lugar de la memoria (donde se encuentra el valor del símbolo, o el proceso con que el símbolo está asociado). Para Newell y Simon la inteligencia consiste en la capacidad de manipular símbolos, es decir, de poder seguir punteros—entes que no nos interesan por lo que son sino por lo que representan. Inteligencia es aptitud para manejar cosas indirectamente: esta indirección es la que constituye el poder propio del pensamiento. Y precisamente, una de las cosas que mejor puede hacer un computador es perseguir una dirección indirecta. Las colecciones de punteros—p direcciones indirectas -debidamente organizadas, dan origen al concepto de *redes semánticas* que ha resultado tremendamente fecundo para ayudar a los practicantes de la IA a desentrañar el inmenso problema de hacer a los computadores capaces de entender y generar lenguajes naturales, como el inglés o el español.

El paradigma computacional se ha mostrado fecundo en el apoyo a la imaginación creadora del científico para concebir importantes leyes que rigen la inteligencia. Los mismos Newell y Simon descubren la ley de proporcionalidad inversa entre potencia y alcance en los métodos del pensamiento. El GPS no fue exitoso como instrumento para replicar la capacidad de resolver problemas que asociamos con el sentido común; pero en cambio, resultó sumamente productivo para revelar los constreñimientos necesarios a que está sometido todo pensamiento. El GPS funciona a base del llamado *análisis de medios y fines*. Consiste en especificar un estado inicial del que se parte y un estado final al que se quiere llegar, en el espacio lógico de los estados posibles en relación con el problema (tales estados deben ser expresables, desde luego, en alguna forma de representación). Operadores o métodos tienen como objetivo reducir las diferencias entre el estado actual de la evolución del problema y el estado final o estado-meta. Parte esencial de la operación del programa es una *tabla de diferencias* donde se determina qué métodos sirven para reducir qué diferencias. Cada operador tiene como requisito para su aplicación el que se den ciertas circunstancias en el estado actual. Si tales circunstancias no se dan, el proceso en curso se interrumpe para dar lugar a la aplicación de uno de los recursos más poderosos del arsenal de los sistemas físicos de símbolos, la *recursión:* el GPS se llama de nuevo, pero ahora para transformar el estado actual en un estado final provisional que sea igual al estado en que alguno de los operadores resulte aplicable, dentro del proceso principal interrumplido.

A primera vista parecería que Newell y Simon hubieran realmente encontrado, en el análisis de fines y medios, la fórmula mágica para la solución de cualquier problema, el verdadero y auténtico solucionador general de problemas, piedra filosofal de la edad electrónica. Pero, lamentablemente, esto es solo apariencia: más bien, su mérito consiste en haber descubierto una ley que dice que no es posible construir un programa que sea capaz de resolver todos los problemas, que sólo se puede construir un programa que resuelven bien un grupo pequeño de problemas o un programa que resuelve mal un grupo grande de problemas. Esta es la ley de proporcionalidad entre potencia y alcance, sobre la que volveremos en seguida.

En la Edad Media, y mucho después, científicos dedicados y tecnólogos ilusos se afanan en vano por descubrir una máquina de movimiento perpetuo; la historia de la ciencia no les da crédito por sus esfuerzos, pero en cambio consagra como grandes realizadores a quienes en el SIGLO XIX descubren las leyes de la termodinámica, que no son otra cosa sino una forma letrada de expresar el enunciado negativo: "no es posible construir una máquina de movimiento perpetuo". El intento de construir un solucionador general de problemas es un fallo parecido, pero sus intérpretes son a la vez capaces de formular la ley negativa correspondiente, que es la famosa ley de proporcionalidad a que nos hemos venido refiriendo. Lo que sucede es lo siguiente: el GPS es capaz, en teoría, de resolver cualquier problema, con tal de que le proveamos de una "tabla de diferencias" adecuada para el dominio de que se trate. Surge aquí una importante distinción conceptual, o más bien la formulación dentro del paradigma computacional de una distinción reconocida por el sentido común: la distinción entre habilidad general y conocimiento (piénsese en las pruebas de inteligencia que aplicamos en los exámenes de admisión a nuestras universidades, y en la eterna polémica sobre si debemos medir habilidad general o conocimientos particulares por medio de esos instrumentos). La habilidad es el análisis de fines y medios, mientras que los conocimientos son la tabla de diferencias.

Hoy por hoy la situación está así: los métodos del pensamiento son o métodos débiles, pero aplicables a cualquier dominio (por ejemplo, el método de análisis de fines y medios), o métodos fuertes, pero aplicables sólamente a cierto tipo de problemas. De ahí han surgido dos ramas en la IA: la dedicada a simular la experticia de los expertos humanos, a base de acumulación de reglas que representan conocimiento (herederas de las "tablas de diferencias" del GPS), y la dedicada a desentrañar el misterio del sentido común, una habilidad que aparentemente es completamente no especializada. Sin embargo, existe la grave sospecha, que estaría respaldada en la ley de proporcionalidad entre potencia y alcance, de que el sentido común como tal no existe: que lo que así llamamos no es otra cosa que una acumulación, en una memoria sumamente flexible y con poderes de recuperación de información excelentes, de experticias superficiales sobre miles de campos especializados, como relaciones humanas, física ingenua, y otras muchas dimensiones en que se desenvuelve el ser humano desde su infancia. Sabremos si eso es así cuando podamos replicar en un programa suficientemente versátil esa capacidad maravillosa del sentido común que pareciera no habernos costado nada, pero que por supuesto es resultado de una evolución milenaria y del aprendizaje espontáneo de muchos años de observación, ensayo y error, y reflexión por parte de los especimenes jóvenes del género sapiens.

Por el momento, cabe decir que la investigación en IA ha comenzado a dar frutos tecnologicos importantes en el otro extremo del espectro de la ley de proporcionalidad. Los sistemas expertos requieren muchos conocimientos sobre un dominio reducido (mínimo alcance), pero con ello obtienen máxima potencia, resultan por lo menos tan efectivos como los expertos humanos correspondientes, a los cuales tratan de imitar y muchas veces superan. Tomemos por ejemplo el sistema FALCON, desarrollado en la Universidad de Delaware por Chester y Lamb. Tiene por objeto el supervisar el funcionamiento de una planta de productos químicos, velando porque todos los indicadores muestren valores compatibles con la seguridad de la planta, y dando la alarma, y un diagnóstico del problema, cuando algo empieza a andar mal. Para realizar su función el sistema usa un intérprete que controla los procesos de observación y deducción, una serie de reglas de inferencia y una amplia base de conocimientos obtenidos con paciencia de los expertos humanos que han realizado el trabajo de supervisión hasta ahora. La existencia de programas como estos ha motivado a filósofos como James H. Moor a considerar seriamente la posibilidad de que los computadores puedan en algún momento sustituir a los seres humanos en la toma de importantes decisiones (MOOR 97).

Los Sistemas Expertos son muy complejos en cuanto al número de reglas que aplican. Son en cambio muy sencillos en la arquitectura de programación. Básicamente todos ellos descansan en alguna forma de "sistema de producción", un método básico que Simon y Newel descubren como operativo en casi todas las actividades intelectuales de los seres humanos (el método tiene un antecedente conceptual en la obra matemática de Post, en la primera parte del siglo).

Un sistema de producción se compone de tres partes: la memoria de largo plazo, donde se encuentran reglas de condición-acción; la memoria de corto plazo, correspondiente a los datos de los sentidos en el caso humano; y el intérprete, que revisa en cada ciclo todas las reglas para ver si alguna combinación de los datos en la memoria de corto plazo corresponde a la condición para la acción de alguna de las reglas —si corresponde, tal regla quedará entre las reglas seleccionadas de ese ciclo. El intérprete aplica luego ciertos criterios de resolución de conflictos para escoger una sola regla del conjunto de reglas seleccionadas; tal regla será la que en definitiva se ejecute, para realizar la acción definida en la misma regla. Normalmente la acción de una regla consistirá en alguna trasformación de los datos existentes en la memoria de corto plazo, por lo que se dice que esa memoria es el canal de comunicación entre las diversas reglas. Es de notar que en ningún caso una regla puede invocar a otra regla, a la manera en que un programa principal en FORTRAN o PASCAL invoca a una subrutina. En realidad, observar el programa no puede darnos ninguna indicación directa sobre cuál regla será la que se disparará en el próximo ciclo: todo depende de las circunstancias en que se encuentre la memoria de corto plazo.

A pesar del éxito que los sistemas expertos han empezado a tener, su rendimiento es todavía demasiado inflexible y compartamentalizado si lo compraramos con sus contrapartes humanas. No obstante, su misma existencia vierte mucha luz sobre los procesos intelectuales del ser humano y sobre la naturaleza de la inteligencia. Significa uno de los primeros grandes éxitos del paradigma computacional, al desentrañar la experiencia

como un conocimiento articulable, frente a posiciones oscurantistas que hacían descansar, por ejemplo, el "ojo clínico" o la capacidad de hacer investigación en un tipo de conocimiento no intelectual y ajeno a la razón. La existencia de sistemas expertos de diagnóstico médico o que permiten diseñar experimentos de biología molecular dan un mentis definitivo a tales pretensiones.

Los esfuerzos de los investigadores por replicar los arcanos mecanismos de la inteligencia han estado siempre mezclados con el intento de descubrir como operan esos mecanismos en la mente del hombre. Por ejemplo, Marvin Minsky (MINSKY 82) analiza la cuestión de si los computadores pueden pensar lógicamente, pero entra de inmediato en un análisis de los mecanismos por medio de los cuales pensamos. Reconoce una inmoderada tendencia a elaborar teorías de la mente que dividen las facultades del cerebro en dos partes de las cuales la izquierda correspondería a la lógica. No le gusta una teoría bipolar porque alienta la insinuación de que cada parte es simple, lo cual no es cierto; ambas son muy complejas. Nos previene además en contra de enfatizar la importancia de la lógica en el pensamiento. Considera deficiente a la lógica por su unidimensionalidad. La lógica se expresa como una sola cadena simplificadora: el hilo del razonamiento. La lógica nos aparta de la representación múltiple, de la redundancia, de la riqueza contextual. La lógica tiende a aislar el texto de su contexto. Y lo que algo significa depende en gran medida de todo lo demás que sabemos, todo lo que decimos supone un contexto. La verdadera riqueza de significado está dada por la multiplicidad de conexiones, y no por la unidimensionalidad que supone la lógica. El científico se siente incómodo ante la dependencia del significado de una expresión sobre el significado de las otras. Por su carácter circular. Prefiere la transmisión lineal de significado desde unas verdades primarias, los axiomas, hacia todo lo demás. Pero eso no destruye el hecho de que en la realidad es así como los humanos adquirimos y manejamos significados.

La teoría de las reglas de significado de Minsky se relaciona cercanamente con la metodología, estimulada en mucho por su trabajo, de las redes semánticas. Pero además ofrece un magnífico complemento a la teoría de los sistemas físicos de símbolos de Newell y Simon. En efecto, es característica de esta última su dependencia sobre el concepto de puntero como esencia de todo simbolismo, y el puntero es equivalente al enlace que conecta a los nodos de una red. En el fondo, los dos enfoques son equivalentes. De ambos se puede criticar, si se quiere, su evidente circularidad, ya que en ellos cada cosa queda explicada por todas las demás, así como en el mundo de Leibniz cada mónada debe contener en sí al universo entero. Otra manera de esgrimir esta crítica es insistir en su solipsismo: no hay manera de salir de la red de significaciones o de símbolos hacia algo más allá de los elementos interconectados, no hay manera de tocar base en el mundo exterior. Pero la verdad es que el solipsismo, no como posición metafísica, sino como actitud metodológica, es la única posición congruente con una actitud científica e incluso con la práxis del ingeniero que aplica la ciencia. Piénsese por ejemplo en el caso de un sistema de computación al que hayamos dotado de conocimientos; siempre dependerá de los órganos de entrada para recibir noticias sobre el mundo, tanto como nosotros debemos también depender de esos oráculos que son los órganos de los sentidos para construir nuestra interpretación del universo exterior. La única garantía que tenemos de su veracidad es el grado de plausibilidad que podamos lograr asignarle gracias a las conexiones recíprocas de todas nuestras significaciones.

Quienes sustenten el nuevo paradigma computacional para la Psicología no pueden soslayar el apremiante problema de la identidad personal o "yo". Minsky tiene una contribución al respecto, que consiste en llamarnos la atención sobre la trampa lógica implícita en la idea de un Agente Unico, verdadero titular de nuestras actividades mentales. La trampa lógica, en la que han caído innumerables filósofos del pasado, como Platón, Aristóteles, Aquino o Descartes, consiste en creer que con dar un nombre interesante a nuestra ignorancia sobre la naturaleza de la mente (como Entendimiento Agente, o Alma Racional, o Substancia Pensante) hemos aclarado en algo los correspondientes fenómenos. Pero hay más: la idea de un Agente Unico detrás de nuestras actividades mentales, de un *homúnculo* escondido en la inteligencia del hombre que es el que *realmente* entiende, es un pecado de petición de principio, de pretender que lo que queremos demostrar está ya demostrado. Porque es obvio que la inteligencia o mente de ese hombrecillo dentro de la mente todavía tendría que ser explicada, si queremos que su presencia contribuya en algo a la aclaración de los fenómenos mentales. Contrariamente a este enfoque, el fundamento de la actitud de Minsky es más bien aceptar, como la mejor lec-

ción que nos haya dado hasta ahora el estudio de los computadores y la investigación en AI, que la mente no es algo simple, no es un Agente Unico, sino por el contrario una construcción sumamente complicada, probablemente el fenómeno de mayor complejidad de todos los que hayan preocupado a la ciencia.

Debemos a Minsky una teoría de la mente basada en la idea de complejidad jerárquica, la *teoría societal de la mente,* presentada por él en su manuscrito inédito del Laboratorio de IA del MIT "Hablemos Claro sobre Epistemología del Desarrollo Nervioso". Básicamente, la teoría supone que el trabajo de la mente es semejante a la labor que realiza un comité, si se quiere, alrededor de una mesa. Cada miembro del comité es experto en algún asunto particular, y dirige el debate cuando el foco de la conversación le corresponde. El conocimiento y la experiencia de cada quien complementa a los de los demás. Viene entonces la pregunta obligada: ¿y qué diremos de la mente de cada uno de los expertos? ¿cómo la explicaremos? La respuesta es que de la misma manera, en el entendido de que cada submente es más simple que su supermente y que en algún momento llegaremos a encontrarnos con mentes menores tan sencillas que ya no necesitaremos invocar la teoría una vez más. Al llegar a ese nivel habremos topado con los *átomos de inteligencia*, es decir, con inteligencias tan simples que no se componen a su vez de partes inteligentes, sino simplemente de circuitos cuyo funcionamiento es elemental (como por ejemplo, el circuito de un termostato, que "siente" el cambio de temperatura y "decide" encender el horno).

Podría pensarse que es parecido postular un homúnculo dentro de la mente para explicar sus poderes y postular un comité de expertos dentro de la mente para lo mismo. Es importante notar la diferencia entre la teoría societal de la mente, que cae de lleno dentro del paradigma computacional por estar basada en la descomponibilidad recursiva de las capacidades intelectuales, y la teoría del Agente Unico o homúnculo. En esta última, lo único que se hace es postular una capacidad con un nuevo nombre para explicar la capacidad inicial; en la teoría societal de la mente, por el contrario, la capacidad se descompone en capacidades cada vez menos importantes, hasta llegar a capacidades primitivas o básicas que ya no necesitan ni pueden recibir, una explicación mentalista. Un mecanismo tan básico como un círculo de retroalimentación, asimilable a un termostáto, será explicable como mecanismo físico, pero ninguna de sus partes, de los conceptos usados en la explicación, pertenecerá ya al ámbito de la Psicología.

No está de más indicar, como una manera de subrayar la fecundidad de las teorías nutridas por el nuevo paradigma, que la teoría societal de la mente —cuyo fin es explicar fenómenos psicológicos— puede ser proyectada "hacia arriba" como forma de explicar fenómenos más amplios que la mente individual. En vez de descomposición aplicaremos ahora composición, pero el procedimiento recursivo será el mismo. Podremos explicar la *inteligencia* colectiva y la inteligencia individual con ayuda de los mismos principios. Pero esto da materia para otro artículo, sobre la posible emergencia en nuestros días de una "Sociología de los Computadores".

Es posible que argumentos como los que he presentado, a pesar de la lucidez que tienen para mí, no convenzan a muchos lectores. Pero espero que estén dispuestos a reconocer que la disponibilidad, dentro del paradigma computacional, de un método para analizar con las herramientas de las ciencias experimentales, temas que habían sido puramente especulativos, ha marcado un cambio fundamental para la Filosofía, la Psicología y las demás ciencias del comportamiento. No podemos ignorar que ahora, por primera vez en la historia, no tenemos que conformarnos con teorizar sobre la composición o funcionamiento de la mente humana. Podemos intentar *construir* una mente. Y aunque no lograramos el ambicioso objetivo de duplicarla, sería muy difícil que tal intento, o sus fracasos, no nos enseñaran muchas cosas importantes sobre el elusivo tema del conocimiento.

REFERENCIAS

Boden 84. Margaret A. Boden. "The social Impact of Thinking Machines". *Futures,* 1984.

Gutiérrez 87. Claudio Gutiérrez y Marlene Castro. *Informática y Sociedad.* San José, Costa Rica, EDUCA, 1987.

Minsky 82. Marvin Minsky. "¿Por qué la Gente Piensa que los Computadores no pueden?" *Al Magazine.* 1982. (Traducido en GUTIERREZ, 87).

Moor 79. Jamen H. Moor, "¿Hay Decisiones que Nunca Deberían Tomar los Computadores?", *Nature and System.* 1987. (traducido en GUTIERREZ 87).

Turkle 84. Sherry Turkle. *The Second Self*, New York, Simon and Schuster, 1984.

Claudio Gutiérrez
1000 San José
Apdo. 3737
Costa Rica