

## UN NUEVO PARADIGMA PARA LAS CIENCIAS DEL CONOCIMIENTO

Lección inaugural de la Universidad de Costa Rica, año lectivo de 1985

**Summary:** *The cognitive sciences (linguistics, cognitive psychology, neural science, epistemology...) developed independently of each other until the middle of this century. Something happened in the 40's that changed that: the formulation by Alan Turing of the universal machine and the invention of the general purpose digital computer. What is common to these two concepts is the capacity of a machine to imitate a machine of the same kind; everything points in the direction of the human mind having the same capacity, which we call knowledge. Thus, the two events taken together constitute a paradigm which has unified and is making much more productive the cognitive sciences.*

**Resumen:** *Las llamadas ciencias del conocimiento (lingüística, psicología del conocimiento, fisiología del cerebro, epistemología, para citar a algunas de ellas) se desarrollaron hasta mediados de este siglo con métodos propios y diferentes entre sí. Algo ocurrió entonces que puede equipararse como un nuevo paradigma (en el sentido de Thomas Kuhn): un avance intelectual (la formulación de la máquina universal por Alan Turing) y una invención tecnológica (el computador digital de propósito general). El elemento común a ambos conceptos es la posibilidad de una máquina para imitar a cualquier otra máquina de su misma especie; todo parece indicar que el cerebro humano tiene esa misma capacidad, a la cual llamamos conocimiento.*

Voy a hablarles de un tema de historia de la ciencia contemporánea. No de toda la ciencia, sino

de las ciencias que tienen por objeto el conocimiento. Algunas de ellas son:

1. La lingüística, que estudia cómo los seres humanos guardamos y comunicamos el conocimiento por medio del lenguaje;
2. La psicología del aprendizaje, que nos dice cómo los seres humanos adquirimos el conocimiento y lo transformamos para hacerlo más poderoso;
3. La psicología cognoscitiva, que trata de aclarar los fenómenos de la percepción, de la atención, y de la memoria, entre otros;
4. La fisiología del cerebro, que analiza los procesos biológicos que sustentan los fenómenos mencionados;
5. La epistemología, parte de la filosofía que tiene que ver con la validez del conocimiento, por ejemplo con el grado de confiabilidad que podemos otorgar a las teorías científicas;
6. La lógica, que estudia las estructuras del conocimiento y los procesos de razonamiento.

Hasta mediados de este siglo, no se pensaba que esas disciplinas tuvieran mucho en común. Cada una de ellas usaba métodos propios muy diferentes entre sí, y era practicada por científicos y filósofos con muy distintos intereses y adiestramiento. Algo ocurrió en la quinta década del siglo que cambió todo eso: un avance intelectual y una invención tecnológica, los cuales asociamos con el nombre de un sabio —Alan Turing— y con un aparato electrónico —el computador digital, respectivamente—. El fenómeno ha demostrado tener la envergadura y trascendencia que asociamos con algunos momen-

tos privilegiados de la historia de la ciencia e identificamos como el surgimiento de un nuevo paradigma científico.

Un paradigma científico, de acuerdo a Thomas Kuhn, es un logro científico de gran importancia que cambia la manera de hacer ciencia de los practicantes de todas las disciplinas afectadas. Un ejemplo es la teoría de Copérnico y su influencia en la manera de practicar la astronomía, o el descubrimiento del oxígeno por Lavoisier y el surgimiento a partir de la consiguiente teoría de la combustión de todo lo que se conoce como química moderna. Otro paradigma de gran trascendencia es la mecánica newtoniana, que logra unir la astronomía y la mecánica terrestre en una superciencia, la física moderna. Todavía otros son la doctrina de la evolución postulada por Darwin y la teoría celular, que cambian en los últimos ciento cincuenta años la manera de practicar las ciencias biológicas. Finalmente, podemos mencionar la obra gigantesca —en nuestro siglo— de los físicos Einstein y Planck, que juntos destronaron al tan exitoso paradigma newtoniano.

La obra cimera del matemático británico Alan Turing consistió en demostrar de manera abstracta la posibilidad de diseñar una máquina autorregulada capaz de imitar el funcionamiento de cualquier otra máquina. Es esta la llamada máquina universal, imposible de construir en la práctica porque requiere una cinta de papel de longitud infinita. Además, la imitación que se alude es independiente del tiempo: puede ser que la máquina universal dure para hacer el trabajo de la máquina imitada varios siglos, en vez de algunas horas o minutos. La invención tecnológica referida es el equivalente posible de la máquina universal: tiene una memoria limitada (en vez de la cinta infinita) y una velocidad de proceso determinada; pero dentro de esos límites puede imitar, como la máquina Turing, cualquier otra máquina —la demostración de Turing nos libró de la pesadilla de tener que oír a vendedores de computadores argumentar “ad nauseam” que el computador de su compañía puede hacer lo que los otros computadores no pueden; como están las cosas sabemos que todo computador de propósito general puede imitar a cualquier otro, y las únicas diferencias en que puede insistir la propaganda se refieren a cantidad de memoria y velocidad de proceso.

¿Cómo se conectan estos acontecimientos con las ciencias del conocimiento? ¿Por qué podemos decir que la máquina universal y el computador de

propósito general proporcionan un paradigma para dichas ciencias? Cuesta poco encontrar la respuesta: ambos, la máquina abstracta y el computador, están dotados de algo que podemos llamar plasticidad intelectual, su capacidad para imitar a cualquier otro ser de su misma especie. Ahora bien, el cerebro humano parece también estar dotado de la misma o muy parecida capacidad: es lo que llamamos conocimiento. En su máxima expresión, el conocimiento es imitación (o emulación) de una mente por otra mente: piensen en lo que ustedes están procurando hacer en este momento, a saber, tratar de reproducir en sus cerebros los procesos mentales por los que el mío va pasando conforme avanza esta charla... Y si el cerebro tiene la capacidad de imitar a la máquina más compleja (otra mente) con más razón posee la de imitar a máquinas más sencillas (objetos físicos o abstractos) que no son máquinas universales.

¿Qué consecuencias se siguen de considerar al cerebro o mente humana como una encarnación, dentro de límites específicos de memoria y velocidad de proceso, de una máquina universal, del mismo modo que lo son los computadores de propósito general? Las consecuencias son extraordinarias: ni más ni menos que unificar las ciencias mencionadas al comienzo, y llevarlas a un grado de productividad jamás alcanzado antes por ellas —exactamente lo que corresponde a una revolución intelectual como las que asociamos con el advenimiento del paradigma newtoniano, o del de Lavoisier, o de Darwin.

Antes de explorar dichas consecuencias, permítanme despejar dos ambigüedades en el concepto de máquina que he venido usando. Cuando decimos que una máquina de Turing es capaz de imitar a cualquier otra máquina debemos suponer que si la máquina imitada no es abstracta sino concreta, la máquina Turing debe no solamente ser realizada de manera física sino también recibir órganos efectores o de sentidos que la capaciten para ejecutar tareas prácticas. Esta complicación técnica, sin embargo, no altera para nada la lógica de la situación, y podemos dejarla de lado en lo que sigue.

La segunda ambigüedad que debe ser despejada tiene que ver con la connotación peyorativa que normalmente asociamos con la palabra máquina. Oponemos los términos “mecánico” y “vital” como lo rígido a lo flexible, lo insensible a lo adaptable, lo grotesco a lo gracioso. Pero nada de eso es válido frente a la plasticidad infinita de una máquina universal, realizada hasta cierto punto en la

plasticidad relativa del computador digital contemporáneo por antonomasia un aparato flexible y adaptable a las circunstancias; por ejemplo, mi sistema de cómputo es capaz de recibir encargos para hacer cuando no lo estoy ocupando, los cuales interrumpe apenas siente que he apretado una tecla para comenzar a darle otra instrucción.

Quizá debiera decir algo, antes de mostrar una aplicación concreta del paradigma computacional, sobre cuál era la situación de las ciencias del conocimiento, o de alguna de ellas, la psicología por ejemplo, antes de su advenimiento. Probablemente todos ustedes habrán oído del célebre experimento de Pavlov a propósito de los reflejos condicionados. Pues bien, tal experimento constituyó un paradigma, de muy dudosa fecundidad por cierto, para la ciencia psicológica: el paradigma conductista, inspirado en experiencias con animales y basado en las categorías de estímulo y respuesta. Lo más importante de ese paradigma es su pretensión de dar fundamento empírico sólido a la disciplina, descartando el método introspectivo con toda su carga de subjetividad y de falta de observabilidad pública. El lado flaco del paradigma es el precio que exige pagar por el logro de tal objetividad: la proscripción del vocabulario mismo con que expresamos la vida mental, los llamados términos intencionales, como "creer", "procurar", "temer", "esperar", etc... que quedan todos substituídos por las respectivas conductas observables. El problema es que distintas "intenciones" se asocian con las mismas conductas, y no hay manera simple de distinguir, por ejemplo, si una amputación es un acto quirúrgico o un acto sádico sin tomar en cuenta las motivaciones de su autor (el contexto no siempre es suficiente para desambiguar: piénsese en las condiciones ascépticas y "hospitalarias" con que se hacían experimentos con seres humanos en la Alemania nazi).

Voy ahora a referirme a una aplicación simple del nuevo paradigma que espero les ilustre sobre lo que su adopción representa mejor que muchas discusiones teóricas. Todos conocen o han oído hablar de computadores que juegan ajedrez. No pueden todavía vencer a los grandes maestros, pero sí a cada uno de nosotros. Si ahora nos preguntamos, mientras presenciábamos imaginariamente una partida en que juega un humano con un computador, cuál irá a ser la próxima movida del humano o del computador, según sea el caso nos encontramos con la siguiente situación:

1. Es igualmente difícil o fácil predecir la movida de cualquiera de los dos contrincantes.
2. La predicción se basará en los mismos principios y circunstancias (básicamente, intenciones de los jugadores y significados, para cualquiera que conozca el juego, de las posiciones del tablero);
3. Ninguna combinación de observaciones sobre estímulos y respuestas sería suficiente para predecir ninguna movida interesante.
4. La predicción deberá basarse en la definición del juego, en la suposición de que el jugador tiene el propósito de ganar, y en la suposición igualmente importante de que el jugador es racional (sabe escoger los medios adecuados para tratar de obtener sus fines).

En relación a la suposición de racionalidad, tendríamos la opción con respecto al computador de decir que el jugador actúa "como si fuera racional". Sin embargo, no parece que intelectualmente ganemos nada con esa maniobra (como tampoco ganaríamos nada con decir que el computador o juega ajedrez, sino solamente actúa "como si jugara"... y el jugador humano reportaría poco consuelo de saber que el computador no lo derrotó sino solo actuó "como si lo derrotara"). En último análisis, debemos reconocer que la capacidad de imitar es esencial para los fenómenos del conocimiento si las mentes de ustedes imitan mis ideas sobre este tema me daré por más que satisfecho. Mucho más productivo es adoptar la estrategia inversa, como lo hace la ciencia cognoscitiva bajo el nuevo paradigma: a saber, suponer que para ser racional (en el sentido indicado) no se necesita ser persona: basta ser un sistema intencional (un sistema capaz de tener creencias, propósitos, planes, etc...).

No tenemos claro en qué consiste para un ser humano tener creencias, propósitos, etc... No obstante, cuando decimos de un computador que pretende fortalecer el flanco del rey, además de saber que tiene en su "mente" lo mismo que tendría un humano que hiciera lo mismo, sabemos también qué quiere decir que tenga ese propósito, puesto que nosotros (algún humano al que podemos interrogar) lo hemos construído y programado; es aquí donde se manifiesta la fecundidad del paradigma computacional que eleva a muy altos niveles las capacidades de predecir y explicar los fenómenos del conocimiento.

Pensemos en las maneras posibles de contestar esta pregunta: "¿Por qué hizo el computador esa

jugada? ” Veamos. Si nos colocamos en la perspectiva del ingeniero electrónico, será complicadísimo contestarla: habría que mencionar innumerables transiciones de dispositivos eléctricos (en cambio, si el computador no hizo ninguna jugada la respuesta a la pregunta “¿por qué no jugó el computador?” podría tal vez ser contestada por el ingeniero de manera muy fácil: por ejemplo, porque la corriente no pasa por este cable roto); esta es una contestación en el nivel físico y a él recurrimos solamente cuando se descompone (enferma) el aparato. El programador del computador podría quizá, con mucho esfuerzo, darnos su propia contestación en el nivel funcional o de diseño, siguiendo el programa paso a paso (pero no podría decir nada sobre el caso del cable roto, como programador); recurrimos al nivel funcional solamente cuando dudamos de la bondad del diseño, por ejemplo si el computador comete un error trivial la explicación del programador sería probablemente sencilla, como “olvidé incluir en el programa que los peones podían comer al paso”. Normalmente, sin embargo, la contestación a la pregunta habrá que darla en un tercer nivel, el nivel intencional: “el computador adelantó el caballo porque cree que el flanco el rey está débil y quiere fortalecerlo”.

El usar el nivel intencional de explicación no implica que los otros no sean posibles; es más, siempre podemos bajar a ellos si las circunstancias lo exigen. Sólo es válido bajo el supuesto de diseño óptimo, que en el caso de los seres humanos muy rara vez cuestionamos (tenemos mucha confianza en la evolución y la selección natural); necesitamos muchísimas muestras de comportamiento aberrante para aceptar la idea de que nuestro interlocutor humano no es racional, es decir, no está en sus cabales; no obstante, a veces lo hacemos, y en ese caso descendemos al nivel de diseño, nos preguntamos por el problema psicológico que la persona tenga, o si el caso es más grave (no nos contesta cuando le hablamos) podemos recurrir también al nivel físico de explicación (presumiblemente ha perdido el oído).

Cuando comenté el paradigma conductista dije que su motivación principal era ofrecer a la conciencia psicológica un fundamento positivo, hacerla tan experimental y firme como la física (el ideal contemporáneo de la unidad de la ciencia exige que en último término y en algún sentido fundamental todas las ciencias puedan ser reducidas a conocimiento físico). Dije también que el precio

que el conductismo tenía que pagar por ese intento era demasiado alto, a saber, negar la validez del uso de vocablos intencionales en las explicaciones psicológicas (¿qué sería de la psicología si le negamos la posibilidad de mencionar en sus leyes cosas como deseos, propósitos, temores, aspiraciones o creencias? . Pero ahora vemos que es posible usar esos vocablos para explicar las actuaciones de un computador, las cuales sabemos que también admiten una explicación que use solamente términos funcionales o físicos pues fue construido por un ingeniero y programado por un programador. ¿Cómo nos escapamos del dilema en que estaba el conductismo? . ¿Cómo ha sido posible este milagro metodológico? . Muy simple: es el concepto de diseño óptimo lo que construye el puente entre los conceptos intencionales o la idea de racionalidad y el terreno firme de las leyes de la física.

La nueva ciencia de la computación, también llamada informática, es la encargada de estudiar estos diseños óptimos, también llamados programas. El paradigma de esta nueva ciencia, fundamentalmente la idea de la máquina universal, es lo que permite en nuestros días considerar a cosas tan dispares como la teoría del cerebro, la psicología del aprendizaje o la inteligencia artificial como aspectos de una sola gran ciencia nueva: la ciencia cognoscitiva o ciencia del conocimiento.

Para terminar, voy a contar y comentar una historia que tal vez ayude a fijar la tesis fundamental de esta charla. En el siglo pasado, mucho antes de la invención de los computadores, un empresario ingenioso anunció poseer una máquina de jugar ajedrez, la cual explotó para su beneficio en varias ferias antes de que se descubriera su truco: un homínulo u hombrecito (un enano), buen jugador de ajedrez, que se escondía dentro de la máquina. El filósofo norteamericano Daniel Dennet aplica esta anécdota en una crítica a la filosofía tradicional que, desde el tiempo de los griegos, ha usado el mismo truco en sus presuntas explicaciones de los fenómenos del conocimiento. En efecto, de modo reiterativo se ha pretendido explicar la capacidad del ser humano para conocer suponiendo algo dentro de ese ser humano que es lo que realmente ejecuta ese conocimiento; por ejemplo, Platón explica el conocimiento como una función de lo que él llama “alma racional”; Santo Tomás de Aquino explica la capacidad racional humana como obra de una instancia interna que denomina “enten-

dimiento agente"; Descartes supone dentro del cuerpo humano, en la glándula pine, algo que califica de "cosa pensante"; ¡evidentemente, todos homúnculos, cada uno de ellos en necesidad de una ulterior explicación que ninguno de estos filósofos siente la necesidad de ofrecer!

Alguien podría sentirse tentado a alegar que el uso de términos intencionales en las teorías psicológicas, se apliquen éstas a mentes o a aparatos, comete el mismo pecado. Por ejemplo, Marvin Minsky, en su "teoría societal de la mente", propone que consideremos a la mente como resultado de la acción coordinada de una sociedad de mentes menores que intercambian mensajes entre sí, cada una de las mentecillas actuando como experto en alguno de los muchos campos que cubre el sentido común humano. ¿No es esta también una ilustración del sofisma del homúnculo? . ¿Nos ofrece alguna salida a esta objeción el paradigma computacional? . Por cierto que sí. La diferencia entre los dos casos es clara y terminante: nosotros sabemos que no hay homúnculos, que cada una de esas

mentes menores puede a su vez explicarse en términos de diseño y, en análisis final, en términos físicos. Pero es útil tomar "préstamos de inteligencia" —que pagaremos después— suponiendo un diseño óptimo del componente que usamos en la explicación de la mente mayor. Los "expertos" de la teoría de Minsky podrán explicarse de la misma manera, como sociedades de mentes todavía menores (menos inteligentes), hasta que eventualmente lleguemos al punto en que una mentecilla muy poco inteligente pueda explicarse como un arco reflejo neuronal sencillo o un circuito digital de carácter básico (habríamos llegado ahí a descubrir los átomos de inteligencia, a saber, pequeñas mentes que son fácilmente explicables sin necesidad de postular componentes internos que todavía posean alguna inteligencia). En ese momento habríamos pagado todas nuestras deudas con la moneda sana de la física, y las disciplinas del conocimiento habrán quedado firmemente asentadas sobre bases estrictamente científicas.