Stoa
Vol. 9, no. 17, 2018, pp. 27–46
ISSN 2007-1868

EL CAMINO HACIA EL AUTÓMATA EMOCIONAL: COMPUTACIÓN AFECTIVA

Rafael Cervera Castellano
Facultad de Filosofía y Ciencias de la Educación
Universidad de Valencia
rafael.cervera@hotmail.com

resumen: La posibilidad de un autómata que pueda pensar y sentir como un humano supone indagar en los problemas fundacionales del movimiento computacional en el siglo xx. La incapacidad de las computadoras bajo la configuración de la Máquina Universal de Turing de realizar computaciones emocionales plantea la hipótesis sobre la existencia de nuevas arquitecturas más adecuadas para el cómputo emocional como las redes neuronales artificiales.

palabras clave: Autómata · máquina de Turing · razonamiento · intuición · emoción · fMRI · mente · cerebro

abstract: The possibility of an automaton that can think and feel like a human supposes to investigate in the foundational problems of the computational movement in the xx century. The inability of computers under the configuration of the Universal Turing Machine to perform emotional computations raises the hypothesis about the existence of new architectures more suitable for emotional computation as artificial neural networks.

keywords: Automaton · Turing Machine · Reasoning · Intuition · Emotion · fMRI · Mind · Brain

Mientras nuestro cerebro sea un arcano, el Universo,
reflejo de su estructura, será también un misterio.
Santiago Ramón y Cajal

1. Pensamiento, computación, e intuición

¿Pueden pensar las máquinas? Alan Turing abría con esta pregunta uno de sus trabajos con mayor impacto en la posteridad filosófica y computacional: "Computing machinery and intelligence" (Turing 1950, p. 433). El tiempo hizo que sus escritos sobre la posibilidad de una máquina con conciencia de sí misma y capacidades pensativas avanzadas tuvieran un impacto desmedido que todavía pervive en nuestros días.

Catorce años antes escribía "On computable numbers with and application to the Enscheidungsproblem" (1936), texto fundacional de la computación moderna y paradigma del mayor ingenio y creatividad lógica expuesta por Turing en los años previos al estallido de la Segunda Guerra Mundial. Principalmente, el texto se proponía abordar la resolución del problema de "decisión" de David Hilbert (1900), el cual era el décimo problema de veintitrés que planteó para ese año y cuya resolución no llegó hasta la segunda mitad del siglo xx para justamente mostrar su no-resolubilidad (Matiyasevich, 1970).

El problema de decisión planteaba la demostración de la existencia de un procedimiento mecánico mediante el cual un cálculo de la lógica de primer orden podría ser un teorema, queriendo decir con esto si sería posible encontrar un modo con el que anticiparse a la resolución de un problema de este orden previniendo su verdad o falsedad.

Turing, y paralelamente Alonzo Church con la realización del cálculo lambda (Church, 1936), abordó el problema estableciendo las bases de lo que sería un "procedimiento" efectivo para establecer la resolución del problema en términos de computabilidad. Ya al inicio de "On computable numbers" define a éstos como "los números reales cuyas expresiones como un decimal son calculables por medios finitos" (Turing 1937, p. 230). Posterior a esta definición, y sin salirse del párrafo, concluirá estableciendo la computabilidad de un número por la capacidad de sus decimales de ser escritos por la máquina en una cinta.

Y no es baladí la importancia de la primera sección de su ensayo, pues en ella están contenidas las bases de todo el movimiento computacional del siglo XX a través del desarrollo de cada vez más avanzadas máquinas "universales". A partir de aquí Turing sigue la estela dejada por Charles Babbage en el siglo XIX con su máquina diferencial para proponer su propia "máquina", la cual era abstracta, pero que a partir de su simbolización podía definir qué era o qué no era efectivo en términos de computación.

De este modo lograba establecer que no debía existir un procedimiento o algoritmo mediante el cual la máquina pudiese determinar que un problema fuese un teorema, al igual que Church también lograra comprobarlo al mismo tiempo, pero por medios distintos. La demostrabilidad de la indemostrabilidad del problema de decisión de Hilbert establecería serias consecuencias mecánicas para lo que un computador pudiese o no pudiese hacer en términos computacionales.

En cierto modo, lo que Turing pretendía quizá demostrar también con este asunto era la incapacidad de su máquina para establecer ciertos tipos de comportamiento que en un humano resultan de lo más sencillos. Previa resolución a un problema matemático, si este no es demasiado complicado, podemos inferir su verdad o falsedad antes de que sea resuelto por la mente humana.

Turing, en su disertación doctoral (1938), describió el razonamiento matemático como dividido entre dos facultades opuestas, pero igualmente necesarias: intuición e ingenio. La primera de estas habilidades cognitivas la definía como cierta capacidad para "hacer juicios espontáneos los cuales no son el resultado de series de razonamientos", mientras que a la segunda "consistía en auxiliar a la intuición a través de disposiciones adecuadas de proposiciones" (Turing 1938, p. 57).

Una Máquina de Turing abstracta dispuesta de una cinta de papel infinita sobre la que escribir cada uno de los números computados, se dice que está realizando un tipo de comportamiento "ingenioso" cuando por medio de un algoritmo procede según lo establecido en éste estableciendo una solución efectiva una vez la máquina se detiene. La máquina se detiene, sí y sólo sí, ésta es capaz de encontrar una solución al problema planteado. En caso contrario, que la máquina no

se detenga nunca corresponde con el tipo de comportamiento que la máquina no sería capaz de simular de ningún modo, la intuición.

Por medio de esta capacidad la máquina podría "saber" cuándo va a parar. Para ello podría encontrarse un algoritmo que dispuesto en ella pudiese determinar si se detendría en algún momento dado o no antes de resolver el problema dispuesto. Este es el famoso problema de la "parada" o detención, el cual establecía esto. No existe un medio bien definido por el que establecer la verdad o la falsedad de un problema aritmético, esto es, que la máquina se detenga en algún punto o continúe imprimiendo números sobre la cinta de manera consecutiva hasta el infinito (de ahí lo de abstracta).

En términos computacionales la cuestión de la intuición como comportamiento procedimental en un computador, analógico o digital, por medio de un algoritmo, se dice que es irresoluble por la misma incapacidad de la máquina de operar de tal modo que lo hace un humano o un matemático. No obstante, se podría aseverar como certera la capacidad "ingeniosa" del computador de, a través de unos pasos bien definidos, resolver un problema aritmético. En este aspecto cabría responder con un Sí a la pregunta de Turing en lo que respecta al pensamiento matemático de la computadora. Pero, ¿sería en algún momento posible un comportamiento intuitivo más allá del estrictamente matemático según lo dispuesto por la Máquina Universal?

2. Computadores electrónicos y cerebros digitales

Si bien es cierto que los ingleses superaron a Enigma a través de la invención de BOMBA, éstos no se quedaron durante el resto de la guerra con el mismo ingenio mecánico para encriptar sus mensajes, sino que dispusieron de la Lorenz SZ42 para asegurar sus comunicaciones (Severance, 2012).

La máquina que en un inicio diseñaron Turing y compañía para combatir a los alemanes era una Máquina Universal o Máquina Universal de Turing, queriendo decir por "universal" su capacidad para simular cualquier otra Máquina de Turing (MT). Desde la perspectiva moderna, BOMBA es el hardware y la MT es el software.

Lamentablemente, para seguir avanzando en la guerra con su ventaja estratégica los ingleses necesitaban en Blechley Park de una máquina más potente capaz de romper los mensajes enemigos tan rápido

como lo hacía BOMBA. Para ello concibieron el diseño de COLOSSUS, el computador universal electrónico.

Puesto en funcionamiento a finales de 1943, COLOSSUS empezó a "romper" los mensajes de Tunny, nombre empleado por los ingleses para referirse a las comunicaciones alemanas, hasta el fin de la guerra (Copeland, 2012). Aunque, a pesar del refinamiento electrónico y lo avanzado de su construcción, estos ingenios seguían sirviendo para un propósito específico, descifrar los mensajes encriptados de la máquina Lorenz. Con posteridad al fin de la guerra se seguiría por la senda computacional diseñando más y más computadores avanzados que pudieran servir con un propósito general tal como funcionan hoy en día.

El hecho de que en la actualidad se pongan tantos esfuerzos en diseñar un computador tal que sea capaz de simular el comportamiento de una persona es consecuencia directa de las investigaciones de Turing y las conclusiones a las que llegó con el desarrollo de los computadores electrónicos. Simular el comportamiento de una persona no es otra cosa que imitarla, esto es, entrar en el juego de la imitación.

Desde este punto de vista, BOMBA y COLOSSUS lo que hacían era imitar de manera bastante sofisticada el comportamiento de Enigma y Lorenz, respectivamente. A través de este "juego" se conseguía saber en todo momento cuál era la configuración exacta de las máquinas de encriptación pudiendo predecir su comportamiento y la situación exacta de sus mecanismos en tiempo real.

En el programa de radio de la BBC de 1951, dando por finalizada la Segunda Guerra Mundial y el desarrollo del computador electrónico-digital, Turing lanzaba al público una sencilla afirmación, que podría pasar fácilmente por alto para la mayoría de oyentes, pero cuyas implicaciones establecerían las bases para la consolidación de la inteligencia artificial moderna (Copeland, 2004):

> Los computadores digitales a menudo han sido descritos como cerebros mecánicos. [. . .] Creo que pueden ser usados de un modo que sea apropiado llamarlos cerebros. Y debería también decir que si cualquier máquina puede ser descrita apropiadamente como un cerebro, entonces cualquier computador digital puede ser descrito de tal modo. (p.482).

Turing concibe esta definición a partir de las funciones que un computador universal realizaba de manera común según su experien-

cia. Más allá del desciframiento de códigos pensó que este tipo de computadores electrónicos serían capaces de realizar tareas específicas según cómo fueran programados y su estructura interna. El computador, como máquina de cálculo, sería capaz de imitar el funcionamiento de cualquier otro ingenio, ya fuera artificial o biológico, si estaba diseñado para tal propósito y cumplía con los requisitos necesarios para ello (Copeland, 2004).

Principalmente, Turing asume que si el computador está debidamente programado debería ser capaz de "comportarse" como un cerebro, tener pensamientos humanos. Para realizar esta tarea sería también necesario, como lo fue con BOMBA y COLOSSUS, poder predecir su comportamiento para tener un sentido completo de cómo funciona el ingenio que se pretende imitar. Al igual sería necesario que el computador tuviera una memoria lo bastante grande que pudiese almacenar los datos de la otra máquina en tiempo real para realizar la predicción de su conducta, y además debería ser lo suficientemente rápida en términos de calculabilidad para realizar esta tarea (Copeland, 2004):

> Si intentamos imitar máquinas todavía más complicadas o cerebros debemos usar computadores más y más grandes para ello. No necesitamos usar necesariamente [computadores] más complicados. Esto puede parecer paradójico, pero la explicación no lo es. La imitación de una máquina por un computador no sólo requiere que debemos haber hecho la computadora, sino que la tenemos que haber programa apropiadamente. (p.483).

Aquí se ve con mayor efervescencia la tesis reinante desde siglos anteriores sobre la condición "mecánica" del cerebro como máquina calculadora capaz de llevar a cabo tantas funciones como caracterizan a una persona singular y que pueden ser extrapoladas a un ingenio artificial.

A partir de las "calculadoras humanas" o calculistas trabajando en distintos sectores de los gobiernos en donde fuera necesario una gran computación de dígitos al instante, se creyó con mayor ímpetu la condición "calculadora" de la mente humana pretendiendo asemejarla con las cualidades ingeniosas de las que hablaba Turing ya en su disertación doctoral (1938).

Adelantándose a su tiempo, como era costumbre, concibió la posibilidad de que un computador pudiese ser "educado" de tal modo que se

comportase como nosotros quisiéramos a partir de un entrenamiento. Turing concibe la analogía con el proceso educacional de una persona a lo largo de su vida. A través de distintos estadios de crecimiento recibimos enseñanzas de un maestro que nos dicta los conocimientos que debemos adquirir. El procedimiento que describe sería el mismo: "Como yo lo veo, este proceso educativo sería en la práctica (algo) esencial para la producción de una máquina razonablemente inteligente dentro de un espacio de tiempo razonablemente corto. La analogía humana sola sugiere esto" (Copeland 2004, p. 473).

La máquina en cuestión dispondría de una memoria o unidad de almacenaje donde guardaría todas sus experiencias a través de tarjetas ordenadas alfabéticamente y por tiempos según el momento en que hayan sido usadas. Se comprende que se trataría de un computador universal dispuesto con una memoria, un sistema de procesamiento, y otro de control, los cuales trabajarían al unísono para resolver los problemas planteados. Turing argumenta a favor de esta tesis en la medida que el "computador" se educa guardando nueva información en su disco duro. Aunque esto no significa necesariamente que la máquina vaya a comportarse en consonancia con lo adquirido y almacenado de un modo diferente ante un input repetido (Copeland, 2004). Este asunto, el que la máquina sea educada y sus respuestas sean formuladas de acuerdo a su poso de datos, se relaciona directamente con el problema de la memoria por un lado, y el del aprendizaje, por otro.

Hilary Putnam abordó la cuestión de la Máquina de Turing y, como consecuencia, la Máquina Universal como versión más avanzada y sofisticada. Su punto de vista es opuesto a la versión más optimista de Turing sobre la "educación" de la máquina a través de un aprendizaje y del uso de su memoria, principalmente porque "la memoria y el aprendizaje no son representados en el modelo de la máquina de Turing como adquisición de nuevos estados, sino como adquisición de nueva información impresa en la cinta de la máquina" (Putnan 1974, p. 133).

A diferencia de cómo operamos los humanos, cualquier dato insertado en la cinta o en la unidad de almacenaje de la máquina no condiciona en absoluto su comportamiento ante un nuevo input, por lo que no aprende nuevos modos de actuación más allá de algún cambio

estructural en su diseño o en la programación de su sistema (Putnam, 1974):

> Entonces si los seres humanos tienen cualquier estado que se asemeja con el estado de una máquina de Turing, aquellos estados deben (1) ser estados en que el humano puede estar en cualquier momento, independientemente del aprendizaje y la memoria, y (2) ser estados totalmente instantáneos de los del ser humano que determinan, junto con el aprendizaje y la memoria, el que será el siguiente estado, así como la especificación total de la presente condición del ser humano (totalmente desde el punto de vista de la teoría psicológica) (p.133).

A este nivel se muestra que tanto un computador como una persona humana mantienen estados mentales a un nivel organizativo distinto estableciéndose la poca semejanza que existe entre los estados de un ingenio de esta clase y los de un humano. En nuestro caso la memoria y el aprendizaje determinan activamente nuestra "identidad" pudiendo responder de manera distinta a nuevos inputs a partir de nuestra experiencia. Se podría decir que nuestros estados mentales gozan de cierta dinamicidad y creatividad. Por otro lado, la máquina de Turing no cambiará en absoluto su identidad por la experiencia recogida, puesto que esta no influirá para nada en su comportamiento futuro (Putnam 1974, p. 133).

Consecuencia de este hecho es la evidencia de la falta de un carácter "productivo" en la máquina. Ned Block y Jerry Fodor (1972) exponen este hecho a través del ejemplo de un autómata probabilista o máquina de Turing, como se ha venido describiendo. Una cualidad asumida entre los humanos es que a través de la experiencia reunida y del modo cómo hemos sido educados en sociedad somos capaces de comprender un conjunto de enunciados proposicionales o experiencias, almacenarlos y establecer respuestas creativas o productivas de acuerdo a ese conocimiento adquirido.

Por cómo está diseñada la MT y la MU (máquina universal), la máquina no es capaz de establecer esta semejanza productiva con el cerebro humano, en contra de lo que propone la teoría de la identidad del estado funcional. Esta teoría clama por identificar una correlación entre los estados funcionales de un computador y los estados funcionales de un organismo a nivel cognitivo con independencia del modo como

esté estructurado su hardware o el material con el que esté hecho. Es por esto que Block y Fodor establecen que "si los inputs y los outputs de un organismo son recursivamente enumerables, entonces se sigue que existe una máquina de Turing capaz de simular el organismo" (Block y Fodor 1972, p. 163).

Pero ya se ha podido comprobar que ni la memoria ni el aprendizaje son determinantes en la máquina con tal de que esta genere un comportamiento productivo y mucho menos que su "identidad" pueda ir evolucionando y cambiando con el tiempo. Igualmente, si como establecieron Alan Turing y Alonzo Church (1938), existen funciones recursivamente no-enumerables en un computador, se asume que habría comportamientos más allá de la deducción aritmética que la máquina no sería capaz de realizar.

No obstante, y paralela a la cuestión de la "productividad", habría que sumar otra objeción a la semejanza de los estados mentales de un humano con los de un computador según lo expuesto por J. A. Fodor y Z. W. Pylyshym (1988), la noción de sistematicidad. Un rasgo característico de las personas es que somos capaces de producir enunciados proposicionales con sentido y organización a partir de la experiencia y la interacción con otras personas.

Cuando proferimos una frase somos capaces no sólo de comprender lo que decimos, sino de proyectar otros significados o frases distintas derivadas de la primera. La acumulación de datos por la MU y la escritura de resultados en la MT no favorecen este tipo de comportamiento siendo bastante difícil, si no imposible, poder desarrollar nuevos outputs en la máquina como consecuencia de un proceso de educación.

Turing nos muestra cómo ciertos comportamientos sí serían posibles en un computador, como el razonamiento ingenioso a través de inferencias realizadas a través de algoritmos codificados en un lenguaje de programación y almacenados en la memoria de la máquina. Pero, aun dando por certeras las previsiones que realizó sobre la existencia de inteligencias artificiales con las que pudiese conversar a través de un lenguaje propositivo (Copeland, 2004), esto no quiere decir que la máquina fuera a "entender" realmente lo que dice. El sistema podría ordenar las frases de manera lo suficientemente creativa para convencernos de que "piensa", pero esto sería falso descubriendo cómo funciona realmente a través de sus circuitos. La inteligencia artificial,

máquina, computador, o como quiera llamarse, posee una habilidad sintáctica, pero no semántica.

Esto es lo que se ha venido a llamar el ejemplo de la caja o habitación china (Searle, 1984) a partir del cual se expone cómo un computador hábilmente diseñado podría establecer respuestas satisfactorias en chino previa a una pregunta en el mismo idioma y siendo que el sistema posee un manual o guía para interpretar los signos chinos de tal manera que pueda establecer respuestas satisfactorias sin entender un solo símbolo de los que dispone "a mano".

En un sentido restringido se podría decir que un computador sí puede pensar en la medida que establece cálculos aritméticos a una velocidad superior a la de un humano. Pero es un hecho que las personas no nos comportamos de esta manera, ni tampoco da la impresión de que poseamos un lenguaje de programación en nuestras cabezas que nos haga funcionar.

Decir con certeza que un computador o autómata dispuesto de un cerebro digital de neuronas "piensa" implica que el agente pensativo es capaz no sólo de razonar y establecer soluciones satisfactorias a ciertos problemas, ya sean discretos o generales, sino que sienta cierta ligazón emocional con lo que hace y dice. Que sea capaz en última instancia de ser un agente artificial moral capaz de tomar decisiones de este tipo tanto de manera razonada como intuitiva.

¿Pero qué nos dicen las ciencias cognitivas de todo eso? ¿Cuál es el sustrato fisiológico del razonamiento deductivo?, ¿y de la intuición? Es más, si las personas no debemos ser vistas como simples calculadoras, sino como seres provistos de una alta capacidad emotiva, ¿cuál es el origen de las decisiones llevadas por una premisa emocional?, ¿y qué papel juega la intuición en todo esto?

3. El cerebro emocional

Tanto una Máquina de Turing abstracta o una Máquina Universal son capaces de trabajar de tal modo que establecen soluciones procedimentalmente satisfactorias según cierto tipo de algoritmo imprimiendo o almacenando información en una cinta o en una unidad de almacenamiento, respectivamente. En "On computable numbers" (1936) se muestra claramente el tipo de operaciones que la máquina es capaz de

computar a través de los, también, llamados números computables. Este comportamiento en la máquina, tildado de "ingenioso", o deductivo, es el que un computador de propósito específico o general es capaz de realizar.

El asunto se complica cuando se pretende simular a través del juego de la imitación un comportamiento intuitivo en el computador. La pregunta por si las máquinas podrían llegar a pensar algún día (Turing, 1950) debe extenderse más allá de lo puramente calculativo en términos aritméticos y considerar los supuestos necesarios para que una máquina pudiese tener comportamientos morales o emocionales ligados de manera amplia con la intuición en las personas.

En la actualidad se han logrado grandes avances en el mapeado del cerebro humano a través de la imagen de resonancia magnética funcional (fMRI) descubriéndose su condición modular. Con esta técnica se lograría "ver" en tiempo real cuáles son los módulos (zonas, secciones) del cerebro implicados en la toma de decisiones morales tanto en un sentido intuitivo como en el razonado ("ingenioso") y el papel de la emoción en todo esto.

¿Pero qué es la intuición? Turing la define como "juicios espontáneos que no son el resultado de pasos conscientes de razonamiento" (Turing 1938, p. 57). Desde el punto de vista psicológico y de la ciencia empírica la intuición se define como "creencias que resultan de un procesamiento psicológico automático, no consciente, y no racional" (Clausen y Levy 2015, p. 171).

Esta idea es aceptada entre el común de investigadores en neuropsicología y ciencias cognitivas, con especial incidencia de aquéllos que defienden la teoría del intuicionismo social (Dellantonio y Job, 2012). Esta teoría establece que las decisiones morales de una persona estarían ampliamente determinadas por su sesgo cultural y social. Razonamiento e intuición no se distinguirían por su *qué* sino por su *cómo*. Ambas facultades del cerebro humano realizan funciones encaminadas a un mismo objetivo, ya sea la resolución de un problema matemático o dar una respuesta clara e inmediata a un problema de carácter social. El razonamiento parece discurrir a través de una serie de pasos razonados mientras que la intuición se los salta dando una respuesta inmediata.

A este respecto se aducen dos posibilidades en ciencias cognitivas. La primera, llamada interpretación continuista, expone que razonamiento e intuición son un mismo proceso, salvo por cómo este se manifiesta según un input sensorial determinado. Mientras que la segunda, la interpretación discontinuista, propone que ambas capacidades pertenecen a dos procesos diferentes (Dellantonio y Job 2012, p. 239). Aunque, de manera independiente a si se trata de uno o varios procesos funcionalmente opuestos, es notable atender a la tesis intuicionista sobre la toma de decisiones morales razonadas o intuitivas.

Esta tesis expondría principalmente que el común de las personas no acostumbramos a establecer procesos de razonamientos bien calculados y "pensados" si debemos interceder en alguna situación a través de un acto moral, y mucho menos razonamos si se trata de decisiones con una alta carga emotiva. A este respecto cabría considerar la preponderancia de las decisiones morales de carácter intuitivo con mayor y especial importancia que el razonamiento intuitivo.

Análisis de fMRI sobre sujetos voluntarios para fases de prueba en cuanto a decisiones de carácter moral o no moral (Greene *et al.*, 2001), se encuentra una especial permanencia de un factor emotivo cuando una acción requiere de un sustrato funcional intuitivo que nos lleve a actuar en situaciones en las que, de cambiarse las variables, no actuaríamos del mismo modo.

Para comprender con mayor profundidad la naturaleza de las decisiones intuitivas y las racionales en el tipo de momentos descrito Greene *et al.* (2001) proponen a un grupo de personas someterse a un experimento en el que tendrán que establecer decisiones de salvaguarda humana en los dilemas del tren y del puente (*trolley dilemma and footbridge dilemma*).

Tanto uno como otro radican en un mismo objetivo, determinar bajo qué condiciones, morales-emocionales, estaríamos dispuestos a sacrificar a una persona para salvar a muchas, ya fuera echándola desde lo alto de una pasarela a las vías del tren (*footbridge dilemma*), o cambiando la dirección del vehículo para que se sacrificara una sola persona en lugar de unas cuantas (*trolley dilemma*). El resultado final siempre es el mismo, alguien debe morir, sólo que a través del escaneo cerebral (fMRI) de los encuestados se conseguirá determinar si estos toman una

decisión bajo un factor emocional o no, y así decidir qué tipo de decisión se trata, si razonada o intuitiva.

Las conclusiones extraídas del estudio muestran cómo hay una clara diferencia emocional entre la elección de un dilema y otro. Siendo el objetivo el mismo, se considera que lanzar a una persona desde un puente o cambiar un interruptor de las vías para cambiar la dirección del tren resultan ambos distintos por la carga emotiva que supone el empujar personalmente a alguien hacia su muerte. Esta diferencia estableció una clasificación en el experimento entre las decisiones intuitivas-personales (con alta carga emotiva), las intuitivas-no/personales (sin emotividad aparente), y las no-morales (Greene *et al.*, 2001).

> Dilemas típicos morales-personales incluyeron una versión del dilema de la pasarela, un caso de robo de los órganos de una persona y un caso de lanzamiento de gente a un bote hundiéndose. [. . .] En cada experimento, nueve participantes respondieron a cada uno de los 60 dilemas mientras experimentaban el escaneo cerebral usando fMRI (p.2106).

Fisiológicamente, el estudio concluye con una exposición de los módulos implicados en la toma de decisiones de carácter moral-personal, y aquellos otros encargados principalmente en los de decisiones morales–no/personales y no-morales. Principalmente se descubre que las zonas activadas durante el escaneo cerebral para las decisiones morales-personales son aquellas estrechamente ligada con la emoción. Mientras que el resto de decisiones se relacionan con los módulos cerebrales destinados a la memoria de trabajo sin ligazón emocional de ningún tipo. Los tres tipos de decisiones son intuitivas y las tres recurren a la memoria implícita del cerebro humano para decidir la mejor opción. Sólo que esta decisión se ve profundamente determinada cuando se trata de decidir a partir de las emociones, eso que justamente nos hace ser personas y no simples calculadoras (Greene *et al.*, 2001, p. 1107).

Este tipo de experimentos de resonancia magnética cerebral aplicados en situaciones donde haya que decidir entre las opciones de conducta suponen una verdadera revelación pues consiguen desenmascarar el sustrato funcional y fisiológico de algo tan sencillo como la intuición humana en relación con la emoción y la moralidad. Más aún,

ayudan a mostrar la actual problemática con los autómatas sobre la posibilidad de que puedan llegar a realizar este tipo de decisiones y los límites funcionales que poseerían para ciertos casos donde se requiera un tipo de discernimiento moral acertado en un instante.

Este tipo de pruebas a través de resonancias magnéticas sobre pacientes que deban elegir entre una situación u otra, en nuestro caso el dilema de la pasarela, ponen de relieve cómo las decisiones de carácter emocional pueden determinar ampliamente nuestro espectro de acción respecto a otras personas que se dejen llevar por decisiones razonadas. En los casos en que las personas del experimento fueron monitoreadas se determinó que las secciones cerebrales implicadas en el procesamiento emocional pertenecían al sistema límbico del cerebro.

Las emociones humanas, descritas desde un punto de vista puramente fisicalista, podrían entenderse como un conjunto de reacciones físico-químicas y/o eléctricas que acontecen en ciertas regiones del cerebro. Estas reacciones son producto de la evolución ante eventos externos que son relevantes para nuestra supervivencia y pueden mantenernos en una situación de alerta para tal fin. Podría definirse a las emociones como un instrumento del cuerpo humano cuyo único fin era garantizar nuestra supervivencia y asegurar el bienestar futuro (Damasio, 2005):

> Las emociones proporcionan un medio natural para que el cerebro y la mente evalúen el ambiente interior y el que rodea al organismo, y para que respondan en consecuencia y de manera adaptativa (p. 56).

Aunque en la actualidad el papel de las emociones es mucho más complejo que hace quince mil años pudiéndose hablar de emociones construidas y otras innatas o básicas en nuestra genética (Prinz, 2004). La teoría sobre que existen una serie de emociones "básicas" ha gozado de especial popularidad en las últimas décadas ganándose cierto consenso entre los investigadores. A este respecto se aduce que existen un total de seis emociones básicas: felicidad, tristeza, miedo, sorpresa, enfado y disgusto. Todas ellas están en la base de nuestro origen evolutivo y su "computación" cerebral permitió que los humanos sobreviviéramos al ambiente (Ekman, 1992a y 1992b).

Con posterioridad se han ido añadiendo más emociones a esta lista, como el orgullo o la vergüenza entre otras, pero lo que importa realmente es cómo está organizado el cerebro humano que permite que podamos experimentar tales sensaciones en nuestro interior (Ekman, 1999).

Las emociones son el resultado de una conjunción extremadamente compleja de distintas redes neuronales repartidas de un modo no-uniforme y descentralizado entre varias secciones con funciones distintas en el cerebro. A este grupo de regiones del cerebro se le llama el sistema límbico y es el encargado de gestionar el procesamiento de las emociones dentro de nosotros mismos de tal modo que podamos reconocer de qué tipo se trata en cada momento (Lotfi y Akbarzadeh-T, 2014; Lotstra, 2002).

Más específicamente, el sistema límbico está formado por la amígdala, el tálamo, el hipotálamo, el hipocampo, el fórnix, el cuerpo mamilar, el bulbo olfativo y el giro cingular. Aunque muchos autores asumen que entre todas estas partes del cerebro las más importantes corresponden al córtex prefrontal, el hipotálamo y la amígdala (Gazzaniga, 2008 y 2009; Damasio, 2005; Banich, 2004; Lotstra, 2002; Fellous, 1999).

El conocimiento sobre las zonas implicadas en el procesamiento de las emociones viene ligado con la creencia actual en que no existen centros localizados en el cerebro humano donde esto sucede. Al margen de lo que podamos pensar, no hay siempre una zona en concreto del cerebro donde se "experimenta" una emoción en concreto, sino que a menudo suelen deslocalizarse entre varias regiones que cumples funciones cognitivas distintas (Gazzaniga, 2008 y 2009; Fellous, 1999 y 2004; Arbib, Fellous, 2004).

El papel de la neuromodulación (Damasio, 2005; Fellous, 1999), es quizá el más importante en la comprensión de las emociones puesto que permite hacernos una idea de cómo trabaja el cerebro en esta cuestión en concreto. Una emoción se comprende que es inconsciente (Damasio, 2005) cuando no interviene un proceso cognitivo consciente, como el razonamiento, para producirla en nuestro interior. A partir de esto, podría establecerse cierta semejanza entre los procesos intuitivos y los razonados que veníamos describiendo.

Atendiendo a la definición de intuición que hemos aportado y los datos que actualmente se están recabando sobre el papel inconsciente de las emociones humanas podría decirse que existe una relación entre este tipo de procesamiento y el modo cómo el cerebro "piensa" las emociones. Siguiendo nuestra argumentación, en última instancia se trataría de establecer cierta semejanza entre un tipo de computación intuitiva-emocional del cerebro humano con el tipo de procesamiento deductivo-razonado de un computador o Máquina de Turing Universal.

4. Computación afectiva

A través de las críticas que se han aducido frente a la Máquina de Turing en relación a la semejanza que se pretendía establecer entre la noción de computador y la de cerebro humano, puede comprenderse la estructura de un computador convencional como la descrita por Turing (1936) o Von Neumann (1945). El computador actual, aunque ha sufrido algunos cambios notables y ha experimentado un gran salto en sofisticación desde mediados del siglo XX, su estructura más básica sigue perviviendo. Tanto Turing como Neumann comprendían que el computador debía de estar formado por una unidad de procesamiento, otra de control y una unidad de memoria. Estas tres partes en conjunto realizaban los procesos computacionales dando un output por cada input introducido.

Analizando, ya no el cerebro humano, sino el de cualquier animal, se ve perfectamente que nuestro órgano pensativo no funciona de la misma manera ni responde a esta configuración. Aunque el hecho de que un computador convencional no sea capaz de neuromodular las emociones como un humano no significa que no sea capaz de un tipo de computación llamada "afectiva" o emocional.

La computación afectiva se entiende como la producción, la imitación o el entendimiento de las emociones humanas con tal de que una máquina sea capaz de reproducirlas y comprenderlas cuando está frente a ellas ante un agente humano u otro artificial (Picard, 2001, 1995). Así se podría comprender a la computación afectiva o emocional como el "reconocimiento, expresión, modelaje, comunicación y respuesta a la emoción" (Picard, 2003, p. 55).

El hecho de que un robot puede pensar o poseer emociones como un humano no es nuevo analizando los distintos autómatas que se han desarrollado a lo largo de la historia humana, su reflejo en obras de ficción o el cine. En la actualidad la tecnología permite que los computadores o sistemas informáticos dispongan de comportamientos emocionales a través de programas de voz o movimientos realistas en autómatas con tal de facilitar la interacción entre humano-máquina. El desarrollo cada vez más avanzado de estos sistemas permite imaginar un futuro en que los computadores y los autómatas no sólo serán capaces de identificar emociones en los demás, sino también de poseerlas y establecer una neuromodulación en sus cerebros digitales (Arbib, Fellous, 2004).

Los autómatas diseñados con este propósito serían capaces de mostrar expresiones faciales complejas en tiempo real como respuesta a un input emocional de un sujeto, artificial u orgánico, el cual la máquina pudiera identificar entre un conjunto de emociones para las que ha sido programado. Esto, unido con la expresión corporal semejante a la de un humano, permitiría una comunicación afectiva entre humano y máquina. No obstante, lograr que un autómata se comunique de este modo a través del "lenguaje" del cuerpo sigue siendo una tarea difícil, por lo que el papel de la neuromodulación emocional aplicada a este tipo de computación es una de las mejores salidas a este problema (Tao, Tan, 2005).

5. Redes neuronales artificiales y emocionales

Resulta agradable cuando un computador o un sistema de voz artificial sonríe alegremente cada vez que lo prendes o te da los buenos días bajo ciertos contextos. Estos sistemas no son más que máquinas de Turing extremadamente sofisticadas cuya estructura computacional ha permanecido casi invariable desde los tiempos que fue concebido, pasando por los prototipos de Von Neumann (1945), hasta nuestros días. Sería notable que una máquina pudiese operar de esa manera todo el tiempo y establecer lazos afectivos con nosotros cuando fuera necesario. Pero, un programa diseñado de tal modo que computara comportamientos emocionales bajo una estructura computacional que no se parece en nada al cerebro humano ni hace uso de los al-

goritmos que describan tal función, no puede entenderse como una computación "real" de las emociones humanas.

Desde mediados del siglo xx con el desarrollo de las primeras redes de procesamiento neuronal (McCulloch y Pitts, 1943; Rosenblatt, 1948.), se ha perseguido la replicación de la estructura neuronal del cerebro humano bajo un sistema artificial con tal de "imitar" los modos de aprendizaje, reconocimiento de patrones y almacenaje de memoria como lo hacemos las personas. Más allá de estas funciones, el cómputo emocional se vio con especial importancia dentro del campo de la inteligencia artificial pretendiendo replicar la biología del cerebro y cómo los distintos módulos se organizan y comunican entre sí para hacer posible las emociones en humanos.

En las dos últimas décadas se han planteado distintas arquitecturas y modelos de redes artificiales con este propósito siendo capaces de aprender emociones según estructuras y algoritmos que copian la interacción química del cerebro (*limbic-based artificial emotional neural network*; Lotfi, Akbarzadeh, 2014), (*brain emotional learning* (BEL); Lotfi y Akbarzadeh-T, 2013a), (*Emotional Artificial Neural Network*; Thenius, Zahadat, Schmickl, 2013), (*emotional back propagation* (EmBP); Khashman, 2010, 2012) o DuoNN (Kashman, 2010). Esta similitud con las redes originales permite que el sistema pueda aprender emociones, reconocerlas y, sobre todo, neuromodularlas con tal de que la máquina no establezca una simple simulación, sino que pueda comprender qué está sintiendo.

El hecho de que la máquina pueda saber lo que está sintiendo se relaciona con capacidades cognitivas tales como la auto-conciencia, la memoria, la atención o la motivación cuyo papel a nivel biológico se está empezando a comprender (Fellous, 2004). El cerebro humano es capaz de dar sentido a lo que sucede en él, por lo que admitir que una máquina realmente "sabe" qué está sintiendo puede resultar una afirmación un tanto apresurada. Pero resulta notable, no sólo el avance en neurociencias sobre el papel de las emociones y la conciencia, sino cómo se está avanzando en la replicación de los sistemas neuronales que hacen posible estas capacidades bajo estructuras artificiales que con el paso del tiempo serán más sofisticadas y permitirán neuromodulaciones más realistas en la persecución de una computación emocional humana en autómatas.

Referencias

Anderson, S. L. y M. Anderson, 2007, "The consequences of human beings creating ethical robots", recuperado de www.reserchgate.net., pp. 1-4.

Arbib, M. A. y J. M. Fellous, 2004, "Emotions: from brain to robot", en *Cognitive Sciences*, vol.8 no.12, pp. 554-61.

Banich, M. T., 2004, *Cognitive Neuroscience and Neuropsichology*, Houghton-Mifflin, Massachusetts.

Block, N. y J. Fodor, 1972, "What psychological states are not", *The Philosophical Review*, no. 81, pp. 159-181.

Church, A., 1936, "A Note on the Entscheidungsproblem", *Association of Symbolic Logic. The Journal of Symbolic Logic*, vol 1, no. 1, pp. 40-41, Doi: 10.2307/2269326.

Clausen, J. y N. Levy, 2015, *Handbook of Neuroethics*, Springer, Nueva York, Doi: 10.1007/978-94-007-4707-4.

Copeland, B. J., 2004, "Colossus: Its Origins and Originators", en *IEEE Computer Society. IEEE Annals of the History of Computing*, vol. 26, no. 4, pp. 38-45.

——, 2004, *The Essential Turing*, Oxford University Press Inc., Nueva York.

Dalgleish, T. y T. Power (Eds.), 1999, *The handbook of cognition and emotion*, John Wiley & Sons, Nueva York.

Damasio, A., 2005, *En busca de Spinoza*, Crítica, Madrid.

Ekman, P., 1992a, "An argument for basic emotions", en *Cognitions and emotion*, no. 6, pp. 169-00.

——, 1992b, "Are there basic emotions?", en *Psichological Review*, vol. 99, no. 3, pp. 550-53

——, 1999, "Basic Emotions", en Dalgleish y Power 1999, pp. 45-60.

Evans.D y P. Cruse (Eds.), 2004, *Emotion, Evolution and Rationality*. Oxford University Press Inc., Nueva York.

Fellous, J. M., 1999, "The Neuromodulatory Basis of Emotion", en *The Neuroscientist*, vol. 5, no. 5, pp. 283-94.

——, 2004, "From Human Emotions to Robot Emotions", en *Cognitive Sciences*, vol.8, no.12.

Fodor, J. A. y Z. W. Phylishy, 1988, "Connectionism and cognitive architecture", en *Cognition*, vol. 28, no. 1-2, pp. 3-71.

Gazzaniga, M. S., 2008, *Human: The Science Behind What Makes Your Brain Unique*, Harper Collins, Nueva York.

——, 2009, *The Cognitive Neurosciences*, A Bradford Book, Londres.

Greene, J. D., R. B. Sommerville, L. E. Nystrom, J. M. Darley y J. D. Cohen, 2001, "An fMRI Investigation of Emotional Engagement in Moral Judgment", en *American Association for the Advancement of Science. SCIENCE JOURNAL*, vol. 293, pp. 2105-2018.

Khashman, A., 2010, "Modeling cognitive and emotional processes: a novel neural network architecture", en *Neural Networks*, vol. 23, no.10, pp. 1155–1163.

——, 2012, "An emotional system with application to blood cell type identification", en *Transactions of the Institute of Measurement and Control*, vol. 34, no. 2–3, pp. 125–47.

McCulloch, W. S. y W. Pitts, 1943, "A logical calculus of ideas immanent in nervous activit", en *Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–33.

Picard, R. W., 1995, "Affective Computing", en *MIT Media Laboratory Perceptual Computing Section Technical Report*, no. 321. Media Lab. Massachusetts Institute of Technology, Cambridge University.

——, 2003, "Affective computing: Challenges", en *International Journal of Human-Computer Studies*, no. 59, pp. 55-64.

Prinz, J., 2004, "Which emotions are basic?", en Evans y Cruse (Eds.) 2004, Doi: 10.1093/acprof:oso/9780198528975.003.0004.

Rosenblatt, F., 1948, "The perceptron: 'A probalistic model for information storage and organization in the brain"', en *Psychological Review*, vol. 65, no. 6, pp. 386–08.

Searle, J., 1984, *Minds, Brains and Science*, Harvard University Press, Massachusetts.

Severance, C., 2012, "Alan Turing and Bletchley Park", en *Computer*, vol. 45, no. 6, pp. 6-8. Doi: 10.1109/MC.2012.197.

Tao J., T. Tan T. y R. W. Picard (eds), 2005, *Affective Computing and Intelligent Interaction*, ACII 2005, *Lecture Notes in Computer Science*, vol. 3784. Springer-Berlín-Heidelberg.

Tao, J. y T. Tan, 2005. "Affective Computing: A Review" en Tao, Tan y Picard 2005, Doi: 10.1007/11573548_125.

Turing, A. M., 1936, "On computable numbers, with an application to the Entscheidungsproblem", en *Proceedings of the London Mathematical Society*, vol. 43, ser. 2, pp. 230-65.

——, 1950, "Computing Machinery and Intelligence", en *Mind* no. 59, pp. 433-60.

——, 1938, *Systems of Logic based on Ordinals*, (Phd dissertation). Presentada en Princeton University y transcrita por Armando B. Matos del Artificial Intelligence and Computer Science Laboratory, Universidade do Porto, Portugal, September 18, 2014.

——, 1947, "Lecture to the London Mathematieal Society", *Charles Babbage Reprint Series for the History of Computing*, vol. 10, The MIT Press, pp. 1-14.