

UNA TEORÍA COGNITIVA DE LAS EMOCIONES

Carlos Herrera Pérez. Glasgow Caledonian University

Quiero hacer un esbozo, casi una pincelada, de una teoría cognitiva de las emociones: cognitiva en cuanto que tener una emoción involucra de alguna manera apreciar, evaluar o conocer algo. Parece aceptado generalmente que las emociones, sin verse reducidas a un evento cognitivo, nos informan, directa o indirectamente, de algo. ¿Qué tipo de conocimiento es éste? No es lo mismo saber el nombre de la capital de Kansas que conocer las leyes de termodinámica, ni resultan similares entender la demostración un teorema y saber que un pié duele, como tampoco sentirse enamorado y saber que uno está equivocado. La apreciación y valoración del objeto de la emoción, sea objeto real o imaginario, físico o formal, está estrechamente ligada a la emoción. Tener miedo ante algo nos indica que ese algo nos parece terrible, amamos lo que nos es amable, odiamos lo digno de odio, etc. Incluso las emociones que parecen no tener objeto, como los estados de ánimo, nos hacen conocer el mundo de otra manera. Quiero esbozar una interpretación de la naturaleza de la relación entre emoción y conocimiento, contraponiéndola a la relación poscognitiva tal como se viene entendiendo desde buena parte de las ciencias cognitivas de corte cartesiano, en particular desde la inteligencia artificial. La motivación es el creciente interés en la construcción de programas informáticos que «reconozcan, expresen e incluso tengan emociones»¹, aun cuando en la mayor parte de los casos el intento se centre en explicar la emoción con la fenomenología ya establecida, más que ver hasta qué punto los términos cibernéticos pueden albergar el concepto de emoción.

1. *Descartes y la inteligencia artificial*

En las discusiones sobre la cuestión mente-cuerpo parece, según los defensores de la IA, que su postura y la de Descartes son antagónicas. Sin embargo, no es éste el caso, y las influencias cartesianas sobre la IA son notorias, aún más en el cognitivismo más 'ortodoxo'. La modernidad, mediante un corte radical entre mente y mundo, funda una nueva concepción del individuo que entiende la inteligencia centrándola en la racionalidad de nuestros pensamientos: sólo cuentan nuestros pensamientos en tanto que desvinculados de uno mismo y del mundo, como razonamiento y conocimiento objetivo. La inteligencia artificial se ha basado siempre en esa tradición, en acometer empresas que no involucren al sujeto que las enfrenta, en particular porque en un ordenador no hay sujeto, al menos en

¹ Una fuente de referencia usual es R. Picard, *Affective Computing*, The MIT Press, Cambridge (Mass.), 1997.

principio. Más concretamente: porque los ordenadores sólo manejan o manipulan información de manera 'objetiva' y desvinculada. Las emociones, tras un análisis superficial siguiendo esa tradición, parece que versan sobre nosotros mismos, sobre el estado de nuestra mente o quizás de nuestro cuerpo, o de cómo nos afecta el mundo, pero en todo caso se centran en nuestra subjetividad y, de alguna manera, atacan esa objetividad racional que es el único camino para un conocimiento firme y un comportamiento seguro.

Descartes no sólo da el pistoletazo de salida para la hegemonía del racionalismo, sino que escinde el ser del hombre en dos esferas bien definidas: cuerpo y alma. Hacerlo metódicamente y con éxito supone que se vayan definiendo las líneas para una nueva concepción de la persona que todavía llega hasta nosotros. Así, para el cognitivismo es fundamental la idea de que existe una división entre lo interno del hombre y lo externo a éste, tal como la describió Descartes. Ahora lo interno está de alguna manera identificado con lo representacional, con una especie de semántica funcional, mientras que lo externo es lo tangible, el tambaleante mundo de la física. Sigue habiendo un dentro y un fuera, con la simple diferencia de que no se otorga un nuevo espacio existencial a lo interno. La mente no puede existir fuera del cuerpo y, sin embargo, se reconstruye el ideal de una mente incorpórea basándose en la siguiente regla de oro a la hora de establecer una explicación cognitiva: los detalles físicos del cuerpo del agente pueden ser ignorados sin ningún riesgo.

Descartes no sólo muestra el camino sino que determina el rumbo a tomar. Dada esta escisión heredada, es curioso encontrar en la literatura sobre inteligencia artificial una nueva versión del «genio maligno» de Descartes². La paradoja del cerebro en una cubeta³ afirma la posibilidad de que YO sea un cerebro instalado en una cubeta como parte de un experimento, no poco maligno por cierto. Este cerebro estaría conectado a una máquina de generación de impulsos neurológicos similares a los que estimulan nuestros cuerpos en un ambiente natural, haciéndonos creer que todo lo que vemos es real a pesar de que está generado por una computadora. Todo lo que percibo y comprendo podría ser una ilusión sobre algo inexistente, porque toda experiencia personal son actos de una mente incorpórea causados simplemente por los eventos del mundo. Se entiende al individuo como metafísicamente independiente de su mundo.

La creencia de que esto es teóricamente factible brota de creer que el mundo no afecta a la persona más que como mundo representado, en parte porque así visto sólo se es persona en tanto que se es representación. Esta representación mental es causada más allá de los límites de nuestro cuerpo, por lo que se admite que sólo podemos tener seguridad de lo que ocurre dentro de esos límites. No vemos el mundo, sino su imagen o efecto en nuestros ojos, su tacto en nuestra piel, su olor y sus ruidos, que percibimos de acuerdo con nuestra naturaleza. Llevando

² R. Descartes, *Meditaciones Metafísicas*, Espasa Calpe, Madrid, 1995.

³ H. Putnam en «Cerebros en una cubeta» hace una descripción clara de la paradoja, reflejada además últimamente en varias películas, de las que la más brillante resulta quizá ser *Matrix*, de los Hermanos Bukowsky. Véase *Reason, Truth, and History*, Cambridge University Press, 1981.

este esquema al límite, no podemos tampoco apreciar nuestras manos o pies, puesto que lo que sentimos son sólo las 'señales' que éstos envían a nuestro cerebro. El cerebro se encarga de 'decodificar' el contenido semántico de esas señales, en un sentido funcional, aunque de alguna manera arbitrario. Es en el cerebro (o incluso en una parte de él que está aún por identificar) dónde reside la mente y la autoconciencia, dónde residimos nosotros. Que el cerebro tome el papel de mente evita una dualidad formal, pero mantiene los esquemas de la dualidad cartesiana. Así, se reedita la dualidad mente-cuerpo, si no en su versión ontológica, ya que mundo externo y cerebro están hechos del mismo material, si al menos en la naturaleza de las conexiones y dependencias entre ellos.

Tanto Descartes como en IA se hace la pregunta: «¿por qué hemos de confiar en las representaciones mentales que experimentamos?» Ahora es la 'bondad' de la naturaleza y de la evolución la que asegura la verdad de las mismas. El argumento cartesiano «Dios es bueno y, por lo tanto, no puede permitir que nos engañemos» se ve transformado ahora en «evolución y selección natural son impecables por ser un mecanismo certero de la naturaleza». Podemos de esta manera confiar en la naturaleza de nuestras percepciones del mundo externo justamente a través de la racionalidad de nuestro mundo interno. No son infalibles, pero si fiables, como confirma nuestra supervivencia en un mundo lleno de depredadores y cataclismos. La finalidad última de nuestro ser y, por tanto, de nuestra mente es la adecuación a ese mundo. De alguna manera fin y ser convergen y se identifican.

Pero el entendernos de esta manera convierte los procesos fisiológicos y psicológicos que llevamos incorporados en mecanismos puramente objetivos, nuestro ojo percibe y codifica información (que bien puede ser imperfecta o incompleta, pero es objetiva), que es tratada virtualmente por nuestro cerebro, y respondemos conforme a nuestro programa de existencia cuando el cerebro envía sus resultados a todos los confines de nuestro cuerpo. Somos simplemente 'una máquina que procesa información'. ¿Qué nos distingue por lo tanto de un ordenador? ¿Dónde entran las emociones?

La descripción de las emociones desde la IA guarda estrecha relación con lo expuesto por Descartes⁴. En el *Tratado de las Pasiones*, los espíritus del cuerpo se veían afectados por su entorno según su naturaleza, por ejemplo ante la presencia de un animal peligroso. Estos espíritus, además de llegar a la glándula pineal para afectar al alma, también afectaban al cuerpo directamente. Claro que para Descartes lo central a la emoción era que afectaran al alma, pero la explicación desde la IA no es distinta.

Las teorías postcognitivas⁵ de la emoción que sirven de apoyo teórico a la IA consideran que, en primer lugar, el sujeto conoce el objeto en cuestión y algunas

⁴ R. Descartes, *Tratado de las Pasiones del alma*, Planeta, Barcelona, 1989.

⁵ A veces se denominan «perceptuales» las teorías que consideran que las emociones no preceden a, o no están basadas en, otros tipos de cognición como creencias y deseos. Las post-cognitivas se llaman simplemente «cognitivas». Véase L. Charland, «Reconciling Cognitive and Perceptual Theories of Emotion: A Representational Proposal» en *Philosophy of Science* 64 (1997), pp. 555-579

propiedades objetivas del mismo, y que más tarde se realiza una evaluación cognitiva de esas características en contraposición con valores u objetivos de alguna manera predeterminados. Dado que en el hombre el procesamiento de información requiere tiempo y que la interacción con nuestro entorno es peligrosa y urgente, la evolución nos ha proporcionado un conjunto de mecanismos de defensa: mecanismos no racionales que resultan ser de alguna manera automáticos. Procesar la información acerca de un animal terrible en el lóbulo cortical, que es donde está nuestra racionalidad, quizá fuera lo más idóneo pero resultaría demasiado lento. Así, procesamos la información en el hipotálamo (nuestro 'órgano' de la emoción), y rápidamente enviamos respuesta a nuestro sistema motor, lo que nos permite escapar a tiempo. Estos mecanismos, los más antiguos en nuestra evolución, llevan consigo una pérdida de nuestra facultad de procesar impecablemente la información, de analizar la situación y tomar la decisión más adecuada. Ver un animal peligroso nos pone en un estado de alerta, incrementando el nivel de adrenalina e invitándonos a poner piés en polvorosa, sin necesidad de paramos a pensar qué hacer. Nuestra racionalidad, el mecanismo principal de comportamiento y apreciación del mundo, se ve así atropellada por las pasiones. Las emociones no son más que la cojera de nuestra racionalidad.

¿Cómo se hace un modelo computacional de las emociones así entendidas? Imaginemos un programa de ajedrez, un programa muy bueno que 'racionalmente', si dispone de mucho tiempo, siempre consiga encontrar la mejor jugada. Imaginemos que las reglas del juego dicen que, hasta que movamos ficha, nuestro contrincante podrá retroceder, arrepentirse y cambiar su último movimiento. Podría ser útil, en ese caso, introducir una orden en el ordenador que diga «si ves que una de tus piezas ataca a la reina del contrincante, tómala con rapidez antes de que se arrepienta de su último movimiento, que seguramente sea un despiste». Claro que quizás sea una trampa, el sacrificio de una valiosa carraza que esconde un jaque mate, una jugada que necesita ser estudiada. Pero no hay tiempo. A eso se le llama «emoción»: el ordenador ha tenido ansiedad.

2. Una receta a base de creencias y deseos

¿Qué estamos presuponiendo en esta descripción? Por lo pronto, que existe un 'programa' óptimo de comportamiento que se ve alterado por las emociones. Presupone que tenemos objetivos claros, como ganar una partida, que tenemos un conocimiento representacional objetivo de lo que nos rodea y que podemos procesar esa información de manera que sepamos qué es lo mejor. Pero, por necesidad, utilizamos a veces otros mecanismos, las respuestas emocionales. Este es el intento de hacer encajar las emociones en una teoría racionalista de creencias y deseos, o conocimiento y objetivos.

Hay resultados empíricos que muestran, sin embargo, que personas con graves deficiencias emotivas tienen a su vez también grandes dificultades en comportarse

de manera inteligente⁶. Quizás las emociones no sean — como tradicionalmente se ha defendido — una perturbación de nuestra racionalidad que nos impida comportarnos con completa eficacia, sino un componente de esa racionalidad. A la vista de esta 'evidencia' empírica, los expertos en Inteligencia Artificial se plantean en los últimos años introducir emociones en ordenadores, haciendo manifiestos en esos intentos tanto su concepción del hombre como de la mente.

Damasio parece dejar claro que es la división entre cerebro y cuerpo la que implica que las emociones sean un edificio más de la arquitectura de nuestro cuerpo ajeno a la mente «no emocional». La división entre unos mecanismos racionales y otros emocionales da entender que podríamos tener los unos sin los otros, porque podríamos tener cerebro sin cuerpo, «la paradoja de la cubeta».

Una explicación neurofisiológica de la mente y de las emociones puede no llegar a ser nunca completa; pero — si abordamos un intento — no podemos restringirnos a las neuronas y a sus conexiones en una parte del cerebro: la mente debe ser un todo en nuestro cuerpo, desde nuestra cabeza a nuestros pies. Las posturas cartesiana y cognitivista, al dividir la persona, colocan a las emociones en terreno de nadie. Por eso en la descripción de las emociones de Descartes el tránsito a través de la glándula pineal es tan fluido. Pero, en el cognitivismo, no hay espíritus del cuerpo para percibir el peligro, así que las emociones son divididas en dos procesos paralelos o secuenciales, de apreciación cognitiva, consciente o inconsciente de la realidad externa, por una parte, y un efecto, por otra, en nuestra interioridad a la vez que en nuestra fisiología.

Quizás sea inútil intentar dar una descripción exacta de lo que el Hombre es términos puramente fisiológicos, pero eso no impide que algunos acercamientos sean más afortunados que otros. El error empezó al llamar «señales» a los procesos del sistema nervioso. Al pinchar un dedo del pie con un alfiler se genera una descarga que altera de neurona en neurona nuestro sistema nervioso hasta alcanzar nuestro cerebro. Llamar a esta descarga «señal que lleva información» da a entender que esa información se puede abstraer y codificar, y que más tarde el cerebro la interpretará. Eso nos hace creer que podemos cortar una pierna e implantar en el muslo un dispositivo que genere ese mismo tipo de señales, pudiendo en la práctica engañar al paciente haciéndole creer que tiene una pierna. La pierna es por lo tanto un accesorio prescindible. Lo mismo acontece con la otra pierna, el tronco, los brazos, e incluso con parte del cerebro. Para la mente no es importante el cuerpo, es simplemente un manantial de señales; ni siquiera es importante la pequeña zona del cerebro donde se genere la experiencia mental, sino sólo la manera en la que el cerebro interpreta esas señales. Pero el cerebro no es una mente, e interpretar es algo que sólo las mentes pueden hacer. Visto así, lo que sentimos no son señales que transportan información hasta el cerebro de, por ejemplo, dolor, sino que sentimos el dolor y lo sentimos en el pie. Y las emociones, por ser su elemento fisiológico imprescindible, no son algo que experimentemos en el cerebro, sino en

⁶ El más influyente de estos trabajos proviene del neurofisiólogo A. Damasio en *Descartes' Error: Emotion, Reason and the Human Brain*, Picador, Macmillan General Books, Londres, 1995.

nuestras entrañas. Las emociones son el paradigma de aquello que no puede ser abstraído a una región de nuestro cerebro.

Además, hay motivos para rechazar que el papel cognitivo de las emociones se pueda reducir a un sistema de creencias y deseos. Las emociones, tanto para Descartes como para algunos cognitivistas, introducen la subjetividad en nuestra mente. De no ser por ellas, el engranaje de nuestra percepción y nuestro cerebro funcionaría según las leyes rigurosas de la naturaleza (o de Dios). Por eso para Descartes lo importante era el control de las pasiones, mientras que en inteligencia artificial se dice que las emociones son un mecanismo creado en nuestra evolución hace mucho tiempo, y que está condenado a desaparecer. Por lo tanto, para la AI, las emociones sólo sirven para perturbar un estado natural ya existente en el hombre, una percepción objetiva a través de nuestros sentidos, codificación y procesamiento objetivo de la información. Nuestros sensores, nuestros ojos u oídos o nariz pueden funcionar erróneamente, pero perciben objetivamente lo que se pone delante. La codificación y el envío de señales al cerebro es también un proceso natural, al igual que el procesamiento racional. Pero este sistema se ve perturbado por las emociones, que son un recurso prehistórico del que todavía no nos hemos deshecho. Las emociones son cognitivas porque implican una apreciación de la realidad y hacen actuar a nuestro organismo en consecuencia. No tenemos tiempo para valorar objetivamente (cosa que podríamos hacer de tenerlo) si el león que va a saltar sobre nosotros es peligroso o no. El proceso inconsciente de esa valoración despierta la emoción de miedo. Más tarde, el haber tenido miedo nos da consciencia del peligro.

Creo haber mostrado cómo está construida esta teoría y qué débiles son algunos de sus pilares. Ahora voy a intentar descubrir algunos problemas que surgen de esa concepción. Una consecuencia de lo expuesto es que la emoción está conectada causalmente con una evaluación de la realidad. Cuando evaluamos algo como peligroso sentimos miedo. Cuando evaluamos algo como doloroso sentimos pena, etc. Concentrándonos en la emoción «miedo», está claro que apreciar algo como peligroso y tener miedo están estrechamente ligados. Para la postura tomada en las teorías post-cognitivas, apreciar algo como peligroso *causa* la emoción «miedo». El peligro causa el miedo en ciertas ocasiones. Como si el peligro fuera una propiedad objetiva de la situación mientras que el miedo fuera una especie de alarma del individuo. Esta explicación— plausible a primera vista— resulta ser de alguna manera circular. En primer lugar, presupone la existencia de un mecanismo de apreciación cognitiva objetiva, consciente o inconsciente, de aquello que es peligroso. Sin embargo, ser peligroso no es una propiedad que algo puede tener como el ser triangular. Podemos distinguir claramente las figuras que son triangulares de las que no lo son, y a pesar de que nos podemos equivocar, nuestra apreciación es objetiva y deberíamos coincidir con cualquier otra persona en las mismas circunstancias. Peligro, sin embargo, no es una propiedad tan claramente delimitada. Ni siquiera la palabra «peligro» está tan claramente definida, ni es tan sencillo su uso en el lenguaje común. Mientras es claro a qué hace referencia «miedo», ¿cuándo es algo peligroso? Una respuesta obvia es cuando supone riesgo

para el sujeto en cuestión. Ahora bien, ¿es «implica riesgo» una respuesta objetiva? Podríamos responder que sí, que conlleva riesgo aquello que estadísticamente ha causado daño en el pasado. Habría que aclarar qué significa «estadísticamente», si es esa estadística está calculada sobre la experiencia personal o sobre la experiencia pública. Creo que esta última opción es absurda, porque el alcance de lo que es arriesgado lo dispondría el instituto nacional de estadística. Si la estadística se basa en experiencias propias, eso explicaría por qué el tabaco, una de las causas de muerte más comunes, no parece peligroso a casi nadie. Al no tener experiencias objetivas cercanas, al ser «peligroso» sólo a largo plazo, el fumar no puede considerarse peligroso. Sin embargo, el miedo a los aviones no puede ser causado por experiencias personales porque yo nunca tuve ningún problema, mi experiencia es impecable y aún así percibo peligro al elevarme en el aire, capto el miedo y el peligro. El peligro y la estadística están relacionados, pero no en la definición de «peligro». Otra opción es decir que algo es peligroso si podemos imaginar una situación en la que cause daño, o sea, de crear de alguna manera una imagen mental de el suceso. Pero la simple imaginación no es suficiente. Puedo imaginar a mi vecino disparándome con un arma; puedo visualizar mi edificio derrumbándose. Pero imaginar no es suficiente: debe ser una imaginación sincera. O sea, imaginar no porque queramos, sino porque sentimos. Pues incluso puedo tener miedo de lo inimaginable, como en el caso de los congéneres de Axtérix y Obélix, que sólo temían ver el cielo caer sobre sus cabezas.

Creo que, llegados hasta aquí, la definición más cercana al uso que hacemos de la palabra peligro sería «aquello de lo que tengo miedo». La versión objetivista de lo peligroso se ha visto relegada a usos en telediarios e informes, pero poco útil para dar cuenta de lo que el miedo es. Por supuesto eso dejaría fuera de lugar que yo hablara sobre el peligro de encontrarse con un oso polar durante una tormenta de nieve, porque no puedo saber si ese oso me daría miedo hasta tener la experiencia. Pero eso no me debe preocupar, porque a pesar de que yo esté convencido de que se oso es peligroso, no debo negar que no tengo idea alguna sobre el tema. Esta nueva definición, además, permite a alguien que nunca haya nadado ni haya visto el mar, quizás un indio de las profundidades del Amazonas en una primera visita al mar, sentir que lanzarse al agua es peligroso. No porque tenga experiencias pasadas, ni porque concluya que el movimiento de las olas podría sumergirle para siempre, sino porque siente miedo. La emoción tiene así un importante aspecto cognitivo, pero no porque refleje que ya hemos apreciado cognitivamente una situación como peligrosa, sino porque es la única manera en la que podemos hacerlo.

La reducción del miedo a la captación del peligro es de alguna manera circular, no porque sea errónea, sino porque es imposible entender la captación del peligro sin el miedo. No es posible dar una definición exacta de «peligro», y, por lo tanto, es aún más implausible pensar que nuestro inconsciente conoce esa definición exacta. Si esa definición existiese, si fuese por ejemplo definida por la estadística, sería fácil reconocer el peligro. En la emoción «echar de menos» el problema es el contrario. Parece fácil describir la cognición que se supone dispara la emoción, sin

embargo no es tan fácil formalizar su apreciación. Según la postura cognitivista se echará de menos al apreciar consciente o inconscientemente la no presencia de alguien querido. ¿Cómo se que alguien no está presente? No es suficiente no verlo porque no vemos a muchas personas conocidas y aún así no somos conscientes de su no presencia. Si la causa de echar de menos es acordarse, recordar o pensar en la persona que es querida, ¿cómo surge ese pensar? Algo, como una foto, una canción, nos hace recordar a la persona amada. Pero lo característico de lo que en portugués se llama «*saudade*» es que todo nos recuerda, que en todo percibimos la ausencia del ser querido. Y es esa justamente la diferencia entre la no-presencia y la ausencia. Ausencia es el objeto que siento en el echar de menos.

A pesar de las descripciones mecanicistas que puedan describir nuestro cuerpo, no somos, por lo tanto, seres que conozcan el mundo objetivamente, a través de sensores mecánicos, codificando y decodificando esa información y tratándola desvinculadamente ateniéndonos a patrones y objetivos preestablecidos. No sólo la emoción lleva consigo un conocimiento sino que el acto mismo de conocer implica una emoción. Cuando vemos algo, no captamos y codificamos su imagen sino que vemos el mundo, y lo vemos tal como nos afecta. La subjetividad no puede comprenderse, por consiguiente, como una falta de objetividad sólo porque en ella perdamos una capacidad para desvincularnos que no es innata. Más bien al contrario, Lo que nos es innato es ser subjetivos, sujetos al mundo, emocionados en él. Ni siquiera podemos entender la objetividad si no es de esta manera, en tanto que somos sujetos emocionados.

Carlos Herrera Pérez
Dept. of Computer Studies
Glasgow Caledonian University
Glasgow, Escocia
C.HERRERA@ GCAL. AC. UK