

## ¿Por qué motivos crearemos máquinas emocionales?\*

Jordi Vallverdú\*\*

Sin duda alguna el siglo XXI está siendo y será el siglo de la Biología. Cuando en el ya remoto año 1953 James Watson y Francis Crick descubrieron estructura de doble hélice del ADN, dieron el pistoletazo de salida para el desciframiento del metafóricamente denominado 'Libro de la Vida', una larga secuencia de seis mil millones de cuatro letras, A, C, G, T, que fue completada apenas pasados cincuenta años. Hoy en día, el estudio del papel de los genes en nuestro comportamiento y funcionamiento ha ido acompañado de avances en los estudios sobre el funcionamiento de la mente. Si bien en la antigüedad el análisis de la mente constituyó el feudo exclusivo de los filósofos, la mente humana es ahora lugar habitual de análisis para psicólogos o neurólogos, quienes comparten con los primeros el interés por los mecanismos que explican la inteligencia humana. Estos mecanismos han sido igualmente tratados de desarrollar e igualar de manera artificial por lógicos, matemáticos e informáticos bajo el amplio proyecto de la Inteligencia Artificial (sistemas expertos, robótica, vida artificial...).

Al fin y al cabo, todos convergen, por vía natural o artificial, en un mismo proyecto: explicar el funcionamiento de la mente humana para solventar sus problemas y aumentar sus capacidades (o desarrollar inteligencias alternativas, a saber, artificiales).

Hasta prácticamente la segunda mitad del siglo XX, todos los investigadores que se acercaban a la mente humana se contraban con muchas dificultades. La primera de ellas, sus propios prejuicios acerca

---

<sup>1\*</sup> Este texto forma parte de las investigaciones del autor como resultados de sus investigaciones para el grupo de investigación TECNOCOG (UAB) sobre Cognición y entornos tecnológicos, [HUM2005-01552], financiado por el MEC (España).

<sup>\*\*</sup> Depto. Filosofía, Universitat Autònoma de Barcelona. E-08193, Bellaterra (BCN), Catalunya - Spain. Tel. 00 34 93 581 16 18. E-mail: jordi.vallverdu@uab.es.

de la pureza de la mente racional (algo prácticamente divino y sin relación con aspectos corporales); la segunda, la comprensión de la estructura neuronal (iniciada por Cajal) y las bases genéticas de la misma; la tercera, la dificultad por estudiar el cerebro en funcionamiento (solventado hoy en día mediante los escáners). Pero no ha sido hasta unas pocas décadas que se realizó un descubrimiento crucial: los procesos racionales están profundamente ligados a procesos emocionales. La mente racional es en realidad una mente emocional.

#### FILOSOFÍA COMPUTACIONAL Y MENTE EXTENDIDA

Existe un fuerte interés por la reflexión sobre el impacto de las ciencias de la computación en el pensar humano. El paradigma contemporáneo ligado a la idea de 'información' (Castells, 2000), ha conllevado un interés por el modo según el cual los datos circulan y crean nuevos significados, formas de trabajo o cambios sociales, provocando incluso la creación de una 'filosofía de la información' (Floridi, 2002). Existen otros nombres para este nuevo pensar sobre los datos de la ciencia y el modo de obtenerlos, procesarlos y analizarlos: filosofía de la computación (Floridi, 2004), ciberfilosofía, filosofía digital (Bynum & Moor, 1998), infoética (Moor 1985) o filosofía computacional de la ciencia (Thagard, 1993). Dejando de lado un análisis clásico de la propia actividad científica, sea a través de las estructuras teóricas, los estudios de caso o conceptos concretos, diversos filósofos computacionales se han interesado por los elementos cognitivos de esta nueva forma de hacer ciencia. La fuerte interrelación entre humanos y entornos tecnológicos ha impulsado fuertemente la idea de la 'mente extensa'. Según esta concepción, nuestras mentes están extendiéndose a través de las herramientas computacionales (Humphreys, 2004), transformando la naturaleza misma de los procesos mentales considerados como un conjunto de operaciones que en estos momentos se encuentran distribuidas en espacios diversos (a caballo entre lo biológico y lo computacional). Nuestro conocimiento es creado a través de una red compleja de elementos que implican

extensiones de nuestros sentidos (robots, prótesis, telescopios, coches, lápices,...) y nuestras mentes (herramientas computacionales).

Existe una larga serie de autores que han contribuido con sus modelos al desarrollo de un marco cognitivo que nos permita entender esta nueva racionalidad a caballo entre lo natural y lo artificial, lo que medio en broma se ha denominado 'el inicio del *wetware*', aunque se lo denomina el 'modelo de la *mente extensa*' (*extended mind*): (Hutchins 1995), (Clark, 2003), (Norman 1990). Asimismo, esta conduce a la idea de la cognición distribuida (Hutchins & Norman 1988) y (Giere 2002), según la cual los procesos cognitivos no tan sólo se producen de forma extensa en un espacio simple que auna mente e instrumentos, sino que este espacio híbrido (mente + instrumentos) se encuentra espacialmente distribuido y compartido por diversos agentes que trabajan de forma conjunta. Disponemos por lo tanto, de un marco cognitivo satisfactorio para el análisis de los aspectos cognitivos de la e-ciencia.

#### COGNICIÓN Y EMOCIÓN: DE LAS NATURALES A LAS SINTÉTICAS

Sin embargo, debemos continuar con nuestro análisis sobre los entornos tecnológicos centrados en un aspecto fundamental de los estudios sobre la cognición contemporánea: las emociones (Ortony et al 1988). Estudios como el de (Norman 2004) consideran aspectos relativos a las emociones. Para realizar un buen diseño cognitivo sobre los procesos racionales, debemos contemplar el papel de las emociones (Dolan 2002). Parte de este cambio radical en el planteamiento del papel de las emociones en los procesos racionales lo debemos a los trabajos de neurofisiólogos como (Damasio 1994) o filósofos como (Pinker 1997, cap. 6, Casacuberta 2000). Debemos pensar que en la base de nuestra capacidad de aprendizaje se encuentra la *empatía emocional*, analizada de forma brillante por Ramachandran (2000) y

Gallese, Kaysers & Rizzolatti (2004). Aprendemos imitando, no sólo los procesos racionales, sino también los sociales y los éticos.<sup>2</sup>

De hecho, en sus investigaciones los científicos están expuestos a lo que Paul (Thagard 2006) ha denominado ‘variables cognitivas calientes’, tales como motivaciones o emociones. Los procesos racionales, por lo tanto, se encuentran condicionados por variables emocionales que condicionan la calidad final de los procesos racionales. Al mismo tiempo, debemos tener en cuenta la naturaleza social de las emociones: estas tienen verdadero sentido en su interacción social, lo que se ha denominado el *social sharing of emotion* (Finkenauer & Rimé 1998), (Rimé et al 1998).

Por otro lado tenemos estudios sobre robots sociables (Breazeal 2002) o computación afectiva (Picard 1997), los cuales se aplican al diseño de robots, máquinas e interfaces que establezcan relaciones emocionales con sus usuarios, bien por la simulación de emociones por parte de tales instrumentos, bien por la permeabilidad emocional de estos a las reacciones de los usuarios o bien por su diseño amable que permita la optimización del trabajo. Incluso se están desarrollando sistemas eficientes de robots sociales, no tanto por lo que respecta a las habilidades sociales en la interacción máquina-humano, sino más bien las existentes entre máquina-máquina (Brooks 1991).

#### ARQUITECTURAS EMOCIONALES.

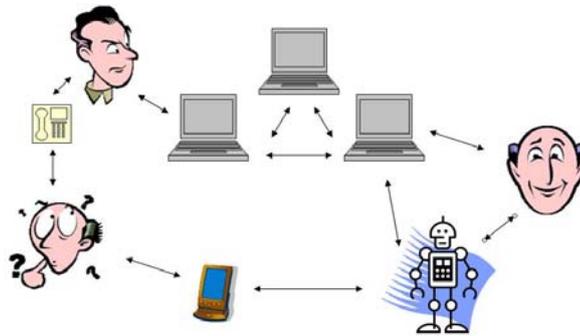
Todo lo anterior nos remite a una aproximación especial al estudio y diseño de un ser eficiente: (1) en primer lugar, la constatación de la ampliación o extensión de lo cognitivo en el ser humano, especialmente en referencia a los entornos computacionales (robótica y software), (2) en segundo lugar, la consideración de los aspectos emocionales de los procesos cognitivos que conducen a acciones racionales; (3) en tercer lugar, el diseño de entornos computacionales que consideren la presencia de aspectos emocionales, que incluyen tanto sistemas de comunicación máquina-máquina (*middleware*), como

---

<sup>2</sup> Sobre este tema, el autor está finalizando la redacción de su innovador ensayo *Una ética de las emociones*, que forma parte del proyecto de investigación “Las fronteras del lenguaje bioético: un nuevo pensar”, financiado por la FVGL.

el uso de interfaces para comunicación humanos-máquinas/comunicación virtual humanos-humanos/manejo de los datos con entornos que lo permitan (*Imaging*)/desarrollo de modelos computacionales sobre los fenómenos.

Veamos un ejemplo gráfico de esta multiplicidad de entornos computacionales e informacionales:



Encontramos una multiplicidad de agentes humanos en contacto con diversos tipos de diseños computacionales con posibilidad de estar conectados en red. Desde un enfoque clásico, se considera el individuo como sujeto epistémico preeminente. Bajo la luz de la teoría de la mente extendida, el sujeto se amplía a su contexto de conocimiento, incluyendo instrumentos. Si tenemos además en cuenta las teorías sobre la cognición distribuida, el número de elementos que colaboran para la creación de conocimiento es todavía mayor. Por último, si consideramos el papel de las emociones en los procesos reacionales y la implementación de las mismas en los entornos artificiales de trabajo (junto con la idea de las emociones distribuidas), tenemos un marco desde el que contemplar y analizar el conocimiento científico contemporáneo: la e-ciencia.

El problema básico en estos momentos, es que no contamos con una arquitectura de las emociones para un proceso cognitivo completo que incluya entes artificiales, sino para elementos fragmentarios de la misma, tales como la relación entre robots y humanos o humanos e interfaces. Las webs semánticas, que permitirán la comunicación

compleja entre máquinas todavía son proyectos de futuro, y en su diseño todavía no se contemplan las emociones de los usuarios que solicitan la información, ni el posible papel de la creación de estados afectivos artificiales en tales sistemas. Un modelo eficiente de racionalidad debería integrar de forma consciente todos estos elementos, permitiendo tanto la consecución de un conocimiento eficiente, como un modelo para la localización de los puntos conflictivos o que conducen a error. Tengamos en cuenta que uno de los problemas de la creciente computerización de la ciencia contemporánea es el de la opacidad que sufren los procesos en red o sometidos al trabajo de sistemas expertos automáticos (Vallverdú 2006). Diseñar entornos más adecuados a las habilidades cognitivas (y por tanto, emocionales) humanas permitiría el diseño de una ciencia extendida que obtuviera grandes ventajas de los entornos computacionales correctamente integrados. Además, una buena Inteligencia Artificial requiere una sólida base emocional, lo que nos lleva inevitablemente al desarrollo de emociones sintéticas.

Si el trabajo en red está creciendo de forma espectacular gracias a mejores y más potentes herramientas computacionales, y si una correcta interacción entre los diversos agentes implica la implementación de elementos emocionales en tales herramientas, resulta obvio que necesitamos de modelos teóricos que contemplen una visión emocional distribuida de la cognición, aplicable a entornos tanto reales como virtuales. Es ya hora de no tan sólo simular, reconocer o reaccionar a emociones, sino de conseguir crearlas en sistemas artificiales, de manera que la cadena de trabajo distribuido sea coherente.

#### CONCLUSIÓN.

A lo largo del artículo se han identificado diversas ideas presentes en la reflexión académica sobre el conocimiento: mente extendida, cognición distribuida, emociones sintéticas, computación afectiva o emociones distribuidas. El presente texto pretende crear un punto de

partida que permita aunar estos conceptos en el diseño de un modelo eficiente de conocimiento extenso, lo que denominado 'arquitecturas emocionales'. Claro está, que con este término no se remite únicamente a los estados afectivos humanos (o sintéticos), sino más bien a un modelo cognitivo humano que contempla el papel de las emociones en los diversos procesos necesarios para la obtención de conocimiento, integrándolos en una arquitectura computacional de niveles diversos (humanos ⇔ máquinas ⇔ máquinas ⇔ humanos). Para ello todavía faltar crear emociones artificiales, e integrar los diversos niveles de interacción entre los agentes en red bajo un único modelo de las emociones que permita un proceso de adquisición del conocimiento coherente y sólido. El protocolo de transmisión entre humanos y máquinas, paradójicamente, (son y) serán las emociones.

#### BIBLIOGRAFÍA

- Arzberger, Peter et al. (2004). "An International Framework to Promote Access to data", *Science*, vol. 303, pp. 1777-1778.
- Breazeal, C. (2002). *Designing Sociable Robots*. Cambridge (MA): MIT Press.
- Brooks, Rodney A. (1991). "Intelligence without representation", *Artificial Intelligence*, vol. 47, pp. 139-159.
- Bynum, T.W. & Moor, J.H. (1998). *The Digital Phoenix. How Computers Are Changing Philosophy*. UK: Blackwell Publishers.
- Casacuberta, David. (2000). *Qué es una emoción*. Barcelona: Crítica.
- Castells, M. (2000). *La era de la información. 3 Vol.*, Madrid: Alianza.
- Clark, A. (2003). *Natural-born cyborgs. Minds, technologies, and the future of human intelligence*, Oxford: Oxford University Press.
- Damasio, R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. London: Harper.
- Dolan, R.J. (2002). "Emotion, Cognition and Behavior", *Science*, vol. 298, pp. 1191-1194.
- Finkenauer, C. & Rimé, B. (1998). "Socially shared emotional experiences vs. emotional experiences kept secret: Differential

- characteristics and consequences". *Journal of Social and Clinical Psychology*, vol. 17, pp. 295-318.
- Floridi, L. (2002). "What is the philosophy of information?", *Metaphilosophy*, vol. 33, n° 1-2, pp. 123-45.
- Floridi, Luciano (ed.) (2004). *Philosophy of Computing and Information*, UK: Blackwell.
- Giere, R. (2002). "Distributed Cognition in Epistemic Cultures", *Philosophy of Science*, vol. 69, pp. 637-644.
- Hey, Tony y Trefethen, Anne E. (2005). "Cyberinfrastructure for e-Science", *Science*, n° 303, USA, pp. 817-822.
- Humphreys, P. (2004). *Extending Ourselves. Computational Science, Empiricism and Scientific Method*. Oxford: Oxford University Press.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge (MA): MIT Press.
- Hutchins, E. & Norman, D. A. (1988). *Distributed cognition in aviation: a concept paper for NASA* (Contract No. NCC 2-591). Department of Cognitive Science. University of California, San Diego.
- Moor, J.H. (1985), "What Is Computer Ethics?" in T. W. Bynum (ed.), *Computers and Ethics*, UK: Blackwell, 263-275. [Published as the October 1985 special issue of *Metaphilosophy*.]
- Norman, D.A. (1990). *La psicología de los objetos cotidianos*. Madrid:Nerea.
- Norman, D.A. (2004). *Emotional design. Why we love (or hate) everyday things*, USA: Basic Books.
- Ortony, A., Clore, G.L., Collins, A., (1988). *The cognitive structure of emotions*, Cambridge (MA): Cambridge University Press.
- Picard, R.W. (1997). *Affective Computing*. Cambridge (MA): MIT Press.
- Pinker, Steven (1997) *How the Mind Works*, USA: W.W. Norton & Co.
- Ramachandran, V.S. (2000) "Mirror Neurons and imitation learning as the driving force behind "the great leap forward" in human evolution", *Edge*, no. 69, May 29.
- Rimé, B. et al. (1998). "Social sharing of emotion: new evidence and new questions". *European Review of Social Psychology*, vol. 9, pp. 145-189.

- Gallese, V., Keysers, C. & Rizzolatti, G, (2004) "A unifying view of the basis of social Cognition", *TRENDS in Cognitive Sciences* Vol.8 No.9 September 2004, pp. 396-403.
- Thagard, P. (1988). *Computational Philosophy of Science*. Cambridge (MA): MIT Press.
- Thagard, Paul. (2006). *Hot Thought: Mechanisms and Applications of Emotional Reason*. Cambridge (MA): MIT Press. [En proceso de edición].
- Vallverdú, Jordi (2006). "Computational epistemology and e-Science", comunicación en *iC&P 2006, Computers and Philosophy, an International Conference*, Laval (Francia), 3-5 de mayo.