

EMOCIONES SINTÉTICAS

David Casacuberta

Jordi Vallverdú

Universidad Autónoma de Barcelona

Resumen: El papel de las emociones en los procesos de toma de decisiones racionales es un hecho demostrado. Al mismo tiempo, el paradigma de la mente extensa resulta útil en el planteamiento de los mecanismos cognitivos implícitos en la ciencia moderna, una e-ciencia altamente computerizada y robotizada, con múltiples elementos procedentes de la Inteligencia Artificial. Por ello, el análisis de modos eficientes de interacción entre seres humanos y entidades artificiales pasa por el estudio de las emociones sintéticas, desde la idea de los robots sociales o los entornos afectivos. El presente artículo analiza estos fenómenos y plantea una pionera propuesta de desarrollo de las investigaciones en esta área.

Palabras clave: emociones sintéticas - e-ciencia - robots - interfaces - mente extensa.

Abstract: The role of emotions in rational decision taking processes is a well established fact. At the same time, the extended mind's paradigm is useful when we analyze the cognitive mechanisms inside contemporary science, an e-science highly technified and computerized, with a deep roots from Artificial Intelligence. Therefore, it is necessary to study the efficient ways of the interactions between human beings and artificial entities, that is, the synthetic emotions: from social robots to affective computing. This paper analyze these topics and offers a new way to develop research in the main area.

Keywords: synthetic emotions - e-science – robots – interfaces - extended mind.

1. Sobre las emociones

1.1. Qué son las emociones y por qué son relevantes para la Inteligencia Artificial

El interés por desarrollar emociones artificiales va mucho más allá de un simple ejercicio filosófico de demostrar que todas las facetas de la vida

humana son mecanizables. Tampoco tiene por objetivo central poder pasar un test de Turing de una forma decente. De hecho, según han mostrado investigaciones recientes, buena parte de las áreas cognitivas que normalmente asociamos con la inteligencia en los seres humanos dependen en buena parte de nuestras emociones. Así pues, incluso tareas de inteligencia artificial de base eminentemente práctica como, por ejemplo, un sistema de *e-learning*¹, pueden beneficiarse con la introducción de emociones sintéticas. (Véase sección 4 de este artículo).

Los seres humanos somos seres fundamentalmente emocionales. Nuestra forma de interactuar con los demás se basa precisamente en la habilidad de comunicar nuestras emociones y de percibir el estado emocional de otros. Extrañamente, esta realidad tan palpable ha sido siempre ninguneada por científicos y buena parte de los humanistas. Históricamente, la cultura occidental se ha caracterizado -con alguna excepción- por mantener una postura desdeñosa en relación a las emociones. Los motivos no son fáciles de establecer. Podrían ser resultado de la influencia del estoicismo (y su idea de que la vida feliz es permanecer impassible, sin que las emociones positivas o negativas nos afecten) en el desarrollo del cristianismo.

También podría ser resultado de cierto pensamiento "patriarcal" que valoraba la razón "masculina" de la acción directa a formas más "femeninas" de ver el mundo apoyadas más en valores comunicativos y sentimientos.

El resultado final es que la mayoría de pensadores y psicólogos han considerado que las emociones eran algo peligroso, incontrolado y no sujeto a la razón. (Véase por ejemplo Lyons (1993) o Damasio 1995). Ello explica en buena parte que hasta bien entrados los años setenta las emociones no fueran objeto de un estudio riguroso desde las ciencias cognitivas y, más en la línea de interés de este artículo- el poco interés que mostraron los desarrolladores de la IA "clásica" por las emociones. También ello explica el mito del "ordenador sin emociones" el razonador puro que tanto pulula en la imaginería de la IA.

Nuestro concepto de razón es demasiado simplista. Hemos opuesto razón y sentimiento cuando en realidad son similares. Nos hemos empeñado en ver a la razón como un producto básicamente discursivo, cuando es sobre todo un vehículo que decide sus rutas centrales a partir de sentimientos.

¹ Por *e-learning* nos referimos a un modelo de formación a distancia que utiliza Internet como herramienta exclusiva de aprendizaje. Esto ha sido llevado a sus extremos dentro del nuevo al paradigma de la e-Ciencia o la Ciencia 2.0., véase Hey & Trefethen (2005) y Schneiderman (2008).

De la misma forma, hemos supuesto que las emociones son el espacio del genio y la locura, la libertad absoluta en la que todo es posible y la razón no tiene nada que decir, cuando en realidad las emociones no son más que las principales garantes de la razón y las que aseguran que no cometemos tonterías. ¡Cuántas estupideces habríamos cometido si no nos hubiera detenido a tiempo el miedo o la vergüenza! ¡Cuántas oportunidades fantásticas nos habríamos perdido si no nos hubiera impulsado la alegría a aprovecharlas! ¿Qué sería de la ciencia, el epítome de la racionalidad, sin esa emoción tan humana de la curiosidad que nos invita a saber más y más de más cosas? Visto desde esta perspectiva, la introducción de las emociones en el mundo de la Inteligencia Artificial no es un lujo o una línea menor de investigación, sino un aspecto central ineludible en la esquematización de comportamientos inteligentes que permitan interactuar con seres humanos.

Así pues, las emociones estaban más allá de la razón, y nos obligaban a tomar decisiones equivocadas. Y cuando no eran ninguneadas, se les hacía un flaco favor, poniéndolas igualmente en el cesto de lo irracional. Románticos y muchos humanistas defensores de las emociones, que argumentan que no sólo de razón pura y de lógica vive el hombre tienden a la misteriosidad y a proclamas irracionales. Cuando los pensadores más humanistas han tratado de recuperar las emociones -como el ya citado movimiento romántico- lo han hecho siempre a expensas de la razón. Las emociones para ellos son algo maravilloso porque nos dan un pensamiento que va más allá de la razón, es misterioso, místico y trasciende la argumentación coherente, dándonos una comprensión intuitiva, directa de los fenómenos. El *dictum* de Pascal de que "El corazón tiene razones que la razón no entiende" es un ejemplo de este "homenaje", aunque calificarlo a uno de irracional y meramente intuitivo no parece ser el mejor favor que a uno le puedan hacer. En el fondo, es darle la razón a la visión tradicional pero desde el otro lado, y el resultado final sigue siendo el mismo, las emociones son elementos irracionales de nuestro pensamiento. Así pues, desde el otro lado de la investigación sobre los fenómenos mentales, no había tampoco ningún incentivo para incluir las emociones en la IA.

Si bien es verdad que en contadas ocasiones las emociones nos han obligado a llevar a cabo algo irracional -piensen en los argumentos de las telenovelas, con protagonistas con las vidas destrozadas por una elección amorosa infortunada- en la inmensa mayoría de situaciones las emociones son eminentemente racionales. ¿No es caso racional huir de un incendio motivados por el miedo? ¿Acaso no es natural y racional sentir repugnancia por un alimento en mal estado y no comerlo? ¿No

resulta útil la sensación de sorpresa que nos obliga a estar más atentos cuando las cosas no van exactamente como estaban planeadas?

Así, el arte, la literatura, los sueños o las emociones se convierten en espacios irreductibles a la razón, inexplicables e intratables por la ciencia, y para ellos es estupendo que sea así.

Sin embargo hallazgos recientes en psicología, especialmente los estudios sobre las personas con daños cerebrales muestran que las emociones son tan racionales como nuestra razón. Antonio Damasio explica en su libro *El Error de Descartes* (Damasio 1995) cómo pacientes con daños en el lóbulo prefrontal no se caracterizaban por su racionalidad, sino por todo lo contrario. En su vida cotidiana mostraban tener claros déficits a la hora de tomar decisiones simples y organizar su vida cotidiana. Así, una persona de profesión oficinista no podía decidir a qué dedicarse en el transcurso de su jornada laboral. Empezaba a redactar una carta y a los dos minutos se levantaba y se ponía a ordenar el archivador para un poco más tarde hacer alguna llamada, pasar rápidamente a leer un informe que abandonaba antes de terminar la primera página para volver a redactar la carta. Sin emoción carecía de motivaciones, y sin motivaciones era incapaz de decidir qué era lo más relevante, por lo que no podía cumplir como se tiene que cumplir con las tareas laborales. La parte más triste de esta historia es que esas personas eran despedidas de su trabajo y no tenían derecho a indemnización, pues según la neurología de entonces no había ninguna relación entre sufrir un daño en el lóbulo prefrontal, en nuestra capacidad de sentir emociones, y en ser eficiente en el trabajo. En una sociedad hiperracional las emociones parecían ser un lujo que no tenía cabida en la toma de decisiones.

Las emociones son sin duda claves en nuestra vida diaria, y es cierto también que importan en nuestro desarrollo personal, cultural y laboral mucho más que el coeficiente de inteligencia. Una persona con una buena capacidad de entender las emociones de los demás -alguien con "inteligencia emocional" según la expresión de Goleman -cfr Goleman (1997)- aunque tenga un CI mediano o incluso tirando a bajo será seguramente una persona con mucho más éxito social, laboral y personal que otra persona con un CI de doscientos pero que no es capaz de interpretar correctamente las emociones de los demás.

Son las emociones y no sólo la consciencia o la razón las que nos convierten verdaderamente en humanos. Más que animales racionales en sentido estricto, somos animales emocionales, y ello no significa que no seamos racionales. Más bien que nuestra razón emana directamente de nuestra emoción. Las emociones son básicas para la inteligencia, pero no

en el sentido pobre de ayuda a triunfar en los negocios: nuestras emociones están en el núcleo mismo de la racionalidad. Sin ellas no podríamos motivarnos y sin motivación no existiría la toma de decisiones, la resolución de problemas y el mismo pensamiento racional. (una buena introducción a la relación entre motivación y emoción es Oatley & Jenkins 1996).

1.2 Características de una emoción que nos interesa capturar en nuestra reconstrucción sintética

Todos sabemos lo que es una emoción intuitivamente, pero lo difícil es explicarlo. En esta sección vamos a intentarlo. No buscamos dar una definición completa, lo que daría para un erudito volumen, sino más bien ofrecer una descripción que aclare alguna de las cosas que decimos de las emociones para entender la forma en que las usamos en inteligencia artificial. En esta sección seguiremos básicamente las ideas planteadas en Casacuberta (2000).

Básicamente, las emociones son procesos psicológicos centrales que reajustan nuestras metas y planes basándose en nuestro propio sistema de valores. Por ejemplo, nuestro plan original de cruzar la calle se puede cambiar si vemos que hay un perro peligroso al otro lado. El miedo es la emoción básica responsable de este cambio. Resumiendo mucho, una emoción es una forma de reaccionar ante un cambio en nuestra jerarquía de planes y objetivos. Primero tenemos que distinguir emoción de un simple acto reflejo -por ejemplo, la acción automática de cerrar los ojos cuando nos llega una luz fuerte, o de una decisión racional "clásica" basada en una argumentación a partir de creencias y deseos conscientes. Por ejemplo, queremos pedirnos un postre de chocolate y nata, pero entonces recordamos que estamos a dieta y que el médico nos ha prohibido estrictamente este tipo de postres, de manera que finalmente pedimos una naranja.

Todo ser vivo tiene una jerarquía de objetivos. En la cúspide están sobrevivir y reproducirse, según vamos bajando tenemos objetivos secundarios también importantes, aunque en menor importancia, como mantener la integridad física o tener un buen trabajo y descendiendo más tenemos objetivos de menor transcendencia como comprarse el nuevo disco de los Arcade Fire o limpiar el coche el fin de semana. Para cada uno de esos objetivos hay una serie de planes específicos que permiten lograrlos.

Normalmente, los objetivos grandes permanecen ocultos, pues no parecen estar amenazados en ningún momento o no requieren una

urgencia inmediata. Afortunadamente, la selección natural no ha puesto en nosotros desalmados instintos de reproducción que nos obliguen a tener ese deseo a cada momento, y confía en que tarde o temprano ya lo haremos. De la misma forma, en un contexto normal no acostumbran a aparecer situaciones en las que nuestra vida esté en peligro.

Sin embargo, imaginemos que la situación cambia. Usted cruzaba la calle con la intención de comprar el último trabajo de los Arcade Fire y de repente un conductor despistado se salta el semáforo en rojo y se dirige directamente contra usted. Es en ese momento en que entra el miedo, usted siente la sensación de miedo y esa misma sensación le lleva a replantearse su jerarquía de objetivos. Su objetivo original de comprar un disco se ve desbordada por la necesidad imperiosa de salvaguardar un objetivo mucho más alto: sobrevivir. Finalmente, su cuerpo activa automáticamente una respuesta, la huida, que le permite -gracias a la generación de adrenalina y a una irrigación y oxigenación extra- saltar y evitar el peligro.

Las emociones son los candidatos perfectos para ocuparse de motivación y toma de decisiones. Implican una reubicación de planes y objetivos y al tener asociadas automáticamente una sensación permiten la motivación directa y el paso a la acción.

Para hacer esta distinción, son necesarias dos cosas: por un lado que parte de la reacción sea automática, una reacción inconsciente generada por la emoción y que es universal. Así, el gesto inconsciente de agacharnos ante el peligro y poner cara de miedo son respuestas, reacciones automáticas a un peligro generadas directamente por una emoción. La reacción de tomar el teléfono y llamar a la policía aunque también está motivada por el miedo no forma parte de la emoción en sentido estricto.

La segunda condición es que la emoción ha de ir siempre acompañada por una sensación, más o menos fuerte, más o menos atenuada, que tiene un efecto final en nuestra decisión y en nuestro comportamiento. Así el miedo viene siempre acompañado por esa sensación característica que nos informa de que sentimos miedo. Lo mismo sucede con la alegría, la tristeza, etc. De esta manera si nos encontramos ante dos personas que huyen de un peligro, una presa del pánico y la otra flemática, podemos suponer que la primera sí está bajo el estado emocional del miedo, ya que siente algo, mientras que la segunda, salvo que sea un actor estupendo, no está bajo la emoción del miedo, ya que no siente nada. Esta característica es central a la hora de distinguir -en filosofía de la mente- entre estados que son realmente emociones de los que no lo son. Sin embargo, la importancia de esta característica en el campo de las emociones sintéticas dependerá de nuestros objetivos: si se trata de un

plan básicamente práctico – queremos desarrollar un sistema que explota las emociones humanas a nivel comunicativo o funcional- en ese caso es totalmente irrelevante la pregunta sobre si el ordenador siente algo. Si por el contrario estamos ante la empresa de crear una máquina que genuinamente siente emociones, necesitaremos haber resuelto primero espinosos problemas en relación a la consciencia, la posibilidad de tener *qualia* y toda una serie de paradojas filosóficas asociadas.

Nuestro gran repertorio emocional se explica por la existencia de una serie de emociones básicas que se combinan entre sí, generando emociones complejas. Es raro que un objeto, persona o situación nos genere una sola emoción, salvo que se trate de una situación muy extrema. Por otro lado, hay eventos que por su propia naturaleza, como una sorpresa agradable, generan varias emociones a la vez. Ello nos permite una vida emotiva muy compleja que explica nuestro arte, literatura, música y también que se hagan necesarias en muchos proyectos ambiciosos de IA.

1.3 Emociones y toma de decisiones

La importancia de las emociones en la toma de decisiones ha sido demostrada experimentalmente en repetidas ocasiones por el ya mencionado Antonio Damasio (Damasio 1995). Este neurólogo de origen portugués ha estudiado detalladamente personas con daños en su lóbulo prefrontal y, por tanto, incapaces de colorear emocionalmente las imágenes mentales que aparecen cuando uno ha de escoger entre diferentes opciones. Los resultados son realmente devastadores para estas personas: incluso en situaciones de laboratorio frente a juegos muy sencillos son incapaces de tomar la decisión correcta y llevan a cabo decisiones que en la vida real les conducirían al desastre. En otro fascinante test en que debían escoger qué personas les parecían de fiar y cuáles no, estas personas, a pesar de tener su memoria y sus habilidades racionales intactas se fiarían absolutamente de todo el mundo, incluso de personas que cualquier personal normal que los viera se aterrorizaría al momento y saldría corriendo de allí, ellos no parecen detectar nada especial.

No hay inteligencia sin motivación, y no hay motivación sin emoción. Todas nuestras decisiones están permeadas por las emociones. Según el modelo de la intervención de las emociones en la toma de decisiones de Damasio, cuando tenemos que decidir una cosa, escoger entre diferentes opciones, esas alternativas se presentan a nuestra consciencia como imágenes mentales. Esas imágenes no son puras informaciones para que

procese nuestra razón, sino que están siempre teñidas de emoción. Nuestros recuerdos y pensamientos están etiquetados con el miedo, la alegría, la tristeza, la repugnancia, etc. que sentimos en aquel momento o que imaginamos que sentiríamos si lo hiciéramos. Además de resultar intuitivamente razonable, hay modelos experimentales en psicología que apuntan en la misma dirección.

Basándose en estas "coloraciones" emocionales, nuestros sistemas emotivo junto al racional hacen un balance y deciden qué opción es la mejor. Si no supiéramos sentir nada especial por ninguna de nuestras posibles decisiones entonces no sabríamos escoger entre ellas, no sabríamos qué es relevante para la acción y que no. No estaríamos motivados. Y tal como hemos dicho al principio, sin motivación no hay inteligencia ni acción posible. Una persona sin esta habilidad es capaz de tomar decisiones extrañas, muchas veces aberrantes, como muestran, por ejemplo, algunos tests para detectar psicópatas.

Decidir no es más que escoger entre un número de opciones. Por ejemplo, supongamos que quiere ir esta noche al teatro. Entre otras cosas debería decidir si tomará el transporte público o se dirigirá allí con su coche. Primero vienen las informaciones: el horario de los trenes, el recuerdo de lo difícil que es encontrar aparcamiento. Las caravanas que se forman al entrar a la gran ciudad. El recuerdo de que la estación de metro más cercana al teatro está a unos diez minutos andando. Una imagen mental de la línea de metro le recuerda que ha de hacer un transbordo para llegar desde la estación de tren a la parada de metro más cercana al teatro... Tradicionalmente, la versión de las decisiones meramente "racionales" pensaba que nuestro cerebro procesaba toda esa información siguiendo reglas formales estrictas y llegaba a una conclusión, y que las emociones no tenían nada que ver.

Sin embargo, ahora sabemos que las emociones tienen mucho que ver. Cuando usted se forma la imagen mental de lo difícil que es aparcar alrededor del teatro, eso no es un dato racional puro, sino que va acompañado por la sensación de agobio que tuvo usted la última vez que se pasó veinte minutos buscando un sitio para aparcar. Igualmente, cuando piensa que se tarda diez minutos en llegar a la estación de metro y en un transbordo, esa imagen viene acompañada por la sensación de cansancio y aburrimiento que le supuso el transporte público la última vez que fue al teatro. Finalmente, la idea de ir al teatro no es una colección de datos objetivos acerca de la obra, sino que va acompañada por la alegría y placer que sintió la última vez que vio a la compañía en cuestión actuando, y la excitación que le ha acompañado últimamente al pensar que podría ver una nueva obra del grupo. Gracias a todas estas

sensaciones, su decisión está finalmente motivada, es capaz de recuperar sólo la información relevante y así decide, en primer lugar, que irá al teatro en lugar de quedarse en casa, pues la obra le interesa mucho, y también decide ir en transporte público, pues el agobio que implica moverse con ferrocarril y metro no es nada comparado a las caravanas de entrada y salida y a la enorme pérdida de tiempo que significa buscar aparcamiento.

Todas estas decisiones no habrían sido posibles sin emociones que le permitieran asignar sensaciones, y por tanto grados de motivación, a cada posible línea de acción.

1.4 Emociones y comunicación

Cuando tenemos una emoción específica, nuestro comportamiento externo y el lenguaje corporal ofrecen diferentes pistas para indicarlo. Lloramos cuando estamos tristes, la voz cambia cuando nos enfadamos, sonreímos cuando estamos contentos, etc. Esta capacidad de comunicarnos mediante las emociones es clave en nuestra vida diaria, y cualquier sistema artificial que quiera asegurar un buen sistema de comunicación entre hombre y máquina tiene que acabar incluyéndolo de alguna forma.

Siguiendo a Casacuberta (2000) podemos establecer tres funciones diferentes de las emociones en los procesos comunicativos:

La función principal es precisamente informar a la comunidad de que estamos bajo cierta emoción, pues ello puede ser útil para el grupo. Por ejemplo, si estamos asustados, es importante que la comunidad lo sepa –a través de nuestra cara de miedo, por ejemplo- pues podría haber un peligro acechando, y todos se van a beneficiar de esa información.

Desde luego, cada uno siente emociones en función de lo que le dicta su cultura: lo que es para nosotros un sabroso bocadillo de jamón puede resultar repugnante para alguien de cultura musulmana. Lo mismo podemos decir acerca de lo que consideramos alegre, triste o agresivo: depende mucho de la cultura. Pero no es menos cierto que la expresión de la emoción es algo universal, independiente de la cultura: todos ponemos las mismas caras de alegría, miedo, tristeza, etc. como han puesto de manifiesto diversos estudios antropológicos.

En buena parte, las emociones se comunican mediante el lenguaje corporal. La expresión facial juega un papel importante en esa comunicación. Diversos estudios desarrollados por el psicólogo Ekman sobre las emociones han mostrado que i) Cada emoción tiene una expresión facial específica y que ii) esa expresión es universal y no

cambia de cultura en cultura. (Cfr. Ekman 1973).

Así, Ekman mostró, por ejemplo, como los habitantes nativos de cierta isla de la Micronesia que no tenían relación con la cultura occidental eran capaces de interpretar qué emoción aparecía expresada en diferentes fotos de ciudadanos estadounidenses que expresaban miedo, tristeza o repugnancia. Igualmente, las expresiones emocionales ofrecidas por los nativos de Micronesia fueron interpretadas de forma correcta por occidentales que no tenían ninguna relación con esta cultura del Pacífico. Como que utilizar el lenguaje habría distorsionado totalmente el ejercicio, Ekman tuvo la inteligencia de utilizar una narración en su lugar. A los nativos se les contaba una historia que estuviera contextualizada a su modo de vida. Así, se les narraba una breve aventura de un hombre que sale a cazar en la espesura y se encuentra con un cerdo salvaje de muy malas pulgas y decidido a atacarlo. En ese momento al nativo se le pedía que pusiera la cara que esa persona pondría al ver el cerdo. El término "miedo" no aparecía en ningún momento de la historia. Se le tomaba una foto y luego se le enseñaban fotos de ciudadanos americanos en diferentes poses. La inmensa mayoría de las veces reconocían la expresión de miedo como la que pondría el personaje en esa situación.

Igualmente, cuando a los ciudadanos americanos se les contaba una historia de miedo -de nuevo sin que el término apareciera- y se les enseñaban diferentes expresiones faciales de los nativos de Micronesia reconocía la gran mayoría de veces la expresión de miedo.

Debido a ello, las emociones son un medio de comunicación muy valioso, al no estar mediado por conceptos, sino que actúa directamente sobre el receptor. Las emociones se pueden comunicar, y entre otras cosas, expresan como una persona se siente por dentro. Cuando uno se siente triste, por ejemplo, y expresa esa emoción a través de lenguaje corporal la gente también se emociona, siente empatía y ayuda a esa persona. De esta forma, las emociones son un fuerte pegamento social en los asuntos humanos, y el mecanismo pasa sin duda por la empatía, sentir lo mismo que siente el otro.

Otra forma en que emociones y comunicación se conectan entre sí es lo que podríamos llamar modulación emocional. Básicamente es la idea es que la forma en que consideramos un evento o persona depende en buena parte de la emoción que tal evento o persona nos genera.

Con Wittgenstein (1922), podríamos afirmar que el mundo del hombre feliz es diferente del mundo del hombre triste. Y no nos referimos exclusivamente al hecho de que una persona deprimida tenga menos ánimos de hacer cosas. Se trata de un efecto más sutil, que hace que la

persona bajo una determinada emoción literalmente perciba el mundo de otra manera. Hemos puesto ya el ejemplo de las personas que, afectadas por una lesión, confían en cualquier individuo, aunque su aspecto sea completamente amenazador y patibulario, pues no son capaces de percibir la amenaza latente.

En otros experimentos psicológicos muy significativos, se ha puesto de manifiesto cómo la manera de interpretar una fotografía o dibujo aparentemente neutro depende de la emoción concreta que una persona está sintiendo. Así, por ejemplo, la imagen de un partido de tenis de aspecto neutro hace que las personas alegres vean gente que se lo pasa bien, sin preocuparse del resultado mientras que las personas deprimidas tienden a ver un perdedor frustrado en el partido con el que automáticamente se identifican. (Cfr. Casacuberta 2000).

Ello también tiene efectos en nuestra actitud comunicativa. Si por una persona sentimos emociones negativas, su mensaje nos llegará muy distorsionado por ellas, y automáticamente generaremos la peor interpretación posible de lo que se nos dice. Tenderemos a estar en contra de las ideas que esa persona nos proponga aunque sean razonables, simplemente porque esa persona "no nos cae bien". Esa emoción no tiene por qué ser necesariamente consciente y asumida. Nuestras emociones también pueden ser inconscientes. Un enfado no detectado puede obligarnos a actuar agresivamente contra una persona, sin ser conscientes de que es el enfado lo que nos hace actuar así. De hecho, es probable que después racionalicemos nuestra actuación, dando una serie de argumentos por las que la persona en cuestión se merecía nuestro trato agresivo. Pero en realidad es nuestro enfado inconsciente, y no las razones que ideamos a posteriori, las que explican nuestro comportamiento agresivo.

2. Cuestiones filosóficas relativas a las emociones (artificiales)

2.1. Los modelos IA y las emociones:

En este apartado consideraremos las diversas aproximaciones a las emociones desde la Inteligencia Artificial (IA), considerada ésta desde una perspectiva amplia que incluya tanto robótica como software, al tiempo que el planteamiento dicotómico de la IA en las líneas *bottom-up* y *top-down*. El primero de ellos, también denominado modelo 'disembodied', se encuentra regido por la *symbol system hypothesis*; mientras que el segundo, también denominado 'embodied' o 'situated', se identifica con la *physical grounding hypothesis*, con relaciones con el conexionismo (la disciplina que trabaja con sistemas simulados de redes neuronales para conseguir reproducir el funcionamiento neuronal del

cerebro humano). Del primero serían ejemplos clásicos autores como (Newell y Simon 1959, 1980), (Simon (1997) o Douglas Lenat (Davis y Lenat 1980), mientras que del segundo brilla por su actividad Rodney Brooks (Brooks 1990, 1991, 1994).

Resumiendo el debate, podemos afirmar que los primeros intentan crear un ente inteligente, una mente artificial, considerada ésta como un tipo de computadora o sistema artificial de alto nivel (prescindiendo de los detalles de implementación del mismo, por considerarlos irrelevantes y de un nivel técnico inferior) que opera con estructuras de símbolos lógicos, mientras que los segundos crean máquinas que imitan determinadas acciones de los seres vivos y consideran que la inteligencia es una propiedad emergente de estos sistemas tras sucesivas construcciones complejas, siguiendo el símil evolutivo. Por estos motivos se los puede agrupar, según (Horvitz 1993), como *formalistas* (los clásicos), o *heuristas* (los nuevos). Los últimos se pueden denominar heuristas pero no en sistemas top-down, sino bottom-up, es decir, 'heuristas de la acción física' (y no simbólica como las heurísticas de los sistemas expertos). De hecho, la IA comenzó durante la década de 1940 desde la visión clásica (todavía existente hoy en día), a la que se opuso hace tan sólo una pocas décadas la visión alternativa. Dentro de esta visión clásica, existiría un período subsimbólico (en la que se trabaja con representaciones numéricas o sub-simbólicas), que abarcaría de 1950 a 1965, mientras que a partir de mediados de los sesenta se inicia un periodo simbólico.

Tras la breve descripción de los dos programas de investigación en IA, el *top-down* y el *bottom-up*, podemos plantearnos cuál ha sido el papel de las emociones en el desarrollo de la IA. Lo cierto es que la aproximación emocional a la IA y los entornos computacionales se produjo inicialmente por cuestiones relativas a la mejora de eficiencia de la actividad humana que trabaja con estos sistemas (Hutchins 1995), (Norman 1999, 2004). Cabe decir que también existen intentos paralelos por descubrir el papel de las emociones en el descubrimiento y evaluación científico (Thagard 2002). Lo cierto, es que a la luz de los estudios sobre el importante papel de las emociones en los procesos racionales, que hemos visto en el apartado 1.1., pronto surgió la necesidad de contemplar la implementación de emociones en sistemas artificiales, con tal de mejorar sus procesos cognitivos y la interacción eficiente de los mismos con los seres humanos (Sloman & Croucher 1981), (Picard 1997). También se ha desarrollado un sistema de modelos artificiales para analizar las interacciones emocionales, denominados *affectons* (Adamatzky 2003), lo que implica una cierta idea de *homeostasis del comportamiento* a partir

de las interacciones emocionales que lo guían (Plutchik 1991). Con ello se busca también la simulación de las transformaciones de las sociedades, en un paradigma de sociedades artificiales que contemple el funcionamiento de las emociones colectivas.

Lo cierto es que este proceso de inserción de lo emocional en lo computacional también respondía a un intento por mejorar nuestra comprensión de la inteligencia humana². Ya en los setenta, y siguiendo la estela creada por Joseph Weizenbaum con ‘Eliza’, Kenneth Colby creó a ‘Parry’, el simulador artificial de paciente paranoico que parecía real para bastantes especialistas (Colby 1981). Las emociones aparecen hoy en día como el caballo de Troya para la comprensión de la racionalidad, de manera que múltiples disciplinas están tratando de aportar luz a esta interesante cuestión (Solomon 1999), (DeLancey 2001)³.

Pasemos ahora a afrontar las equivalencias a los modelos *top-down/bottom-up* en el ámbito de las emociones sintéticas:

2.1.1. Top-down.

Siguiendo la evolución de la IA, los intentos por crear un sistema simbólico, esta vez ligado a las emociones, es algo que se ha producido de forma parcial. CYC es un ejemplo de ello, si bien Douglas Lenat, su creador, considera que no tiene emociones genuinas⁴. Con todo, CYC comprende el significado de las emociones en los entornos donde estas aparecen. Sin ellas, Lenat comprendió que CYC no sería el metasistema experto que perseguía.

Por otro lado, existe una aproximación a lo emocional desde la robótica y los entornos computacionales. Disponemos de un proyecto paradigmático que ha marcado las investigaciones de los últimos años: Kismet. El grupo

² Y de aspectos relacionados con la inteligencia como la intencionalidad o la creatividad. Sobre este último aspecto, tenemos el ejemplo del programa AARON, de Harold Cohen (<http://circa.ucsd.edu/~hcohen/>).

³ (Scheutz 2004) detalla una completa bibliografía histórica de las emociones en la IA.

⁴ <http://www.cyc.com/cyc/technology/halslegacy.html>. CYC es una ontología del conocimiento general. Contiene 100.000 conceptos y 1.000.000 declaraciones y el objetivo de su creador es el de crear una máquina que entienda el mundo en su globalidad del mismo modo que cualquier ser humano normal lo hace. También se supone que será capaz de aprender como lo hace un ser humano, una vez haya adquirido los elementos conceptuales (sintácticos y semánticos) necesarios. Es el sueño de la Inteligencia Artificial. Los estudios de red semántica actuales tan sólo son una aproximación menor al proyecto que persigue CYC. En cualquier caso, no se ha conseguido culminar este proyecto, si bien su desarrollo ha aportado interesantes avances.

de investigación “Robotic Life Group”, del MIT Media Lab, liderado por Cynthia Breazeal está creando robots que puedan interactuar con los seres humanos, por lo que el aspecto emocional de tales máquinas ha sido un punto fundamental en su diseño. La antropomorfización de los robots es un elemento a considerar seriamente en su mayor grado de aceptación (¡al tiempo que de rechazo!), pasando de la consideración del mismo como una herramienta a la del robot como un compañero. Breazeal tras trabajar en el robot Cog, desarrolló a partir de finales de los años noventa el robot Kismet, una cara esquemática (un soporte básico para orejas, ojos, labios y cejas) que si bien en los primeros estadios de la investigación tan sólo era capaz de simular gestos faciales relacionados con las emociones, lo que producía una empatía del usuario humano ante una máquina de este tipo, hoy en día Kismet es capaz de ‘reconocer’ emociones en humanos y reaccionar en consecuencia mediante gestos faciales o emisión de sonidos.

La pregunta clave es ¿tiene Kismet emociones? No, las *simula*. Pero con ello consigue que en la interacción con un ser humano, éste se sienta mejor que al relacionarse con una máquina fría y distante, máxime eficiente. Del mismo modo que Deep Blue, el ordenador de IBM que venció al ajedrez al mejor jugador de todos los tiempos, Gary Kasparov, es imbatible en el ajedrez, no por ello diríamos que es inteligente en el sentido amplio del término... aunque en ajedrez ya no lo ganará humano alguno. Kismet viene a ser su equivalente en el ámbito de la robótica emocional: es un gran logro, pero no tiene emociones genuinas. Su gran aportación es la implementación de habilidades emocionales que permiten una mejor socialización (Bates 1994; Breazeal 1998, 2002; Picard & Klein, 2002), puesto que se ha demostrado que los seres humanos respondemos socialmente a la relación con las computadoras (Reeves & Nass, 1996), a las que consideramos un “embodied computer agent” (Brave, Nass & Hutchinson, 2005). Y la empatía puede considerarse un fenómeno afectivo (Batson et al. 1997). No es raro que a los sistemas conformados por humanos y máquinas que trabajan conjuntamente se los haya denominado ‘sistemas cognitivos conjuntos’ (Hollnagel & Woods, 1983). Es decir, un sistema cognitivo extenso, no exento de problemas como la pérdida de habilidades, la confianza en la bonanza del sistema (que lleva a la complacencia) o la pérdida de adaptabilidad (Parasuraman 1997).

Para la industria, esto no constituye problema alguno. El perro AIBO de la casa SONY (su nombre real es ERS-7M3), simula actitudes propias de una mascota que provocan en sus dueños fantasías con respecto a sus capacidades emocionales (Friedman, Kahn & Hagman, 2003).

En el campo del software, hubo un programa pionero que causó furor a finales de la década de 1960, su nombre: Eliza⁵. (Weizenbaum 1966) fue el creador del primer programa artificial de conversación entre los seres humanos y una computadora. Los modelos computacionales interactivos que contemplan las emociones humanas son una de las líneas de investigación más importantes de nuestros días (Elliot, 1992; Tosa & Nakatsu, 1996). Uno de los aspectos importantes en estos procesos interactivos es el de minimizar o evitar la frustración del usuario, lo que conlleva disminución de capacidades cognitivas de la persona afectada, como en la capacidad de atención, retención memorística, aprendizaje, pensamiento creativo o interacción social educada (Klein, Moon & Picard, 2002).

2.1.2. Bottom-up.

El enfoque *bottom-up* parte de un punto de vista totalmente opuesto al que acabamos de analizar. La idea es que las emociones básicas, dolor, placer, atracción, aversión,... sean partes fundamentales de las máquinas o sistemas creados.

Del mismo modo que no podemos realmente crear máquinas con pensamientos complejos ni tampoco con sistemas de movimiento muy especializados a la par que variados (como en los seres humanos), que es lo que habría intentado un enfoque *top-down* durante décadas, desde la perspectiva *bottom-up* se persigue algo completamente diferente: comenzar por sistemas muy simples que imiten a la naturaleza para ir aumentando paulatinamente su complejidad. Es lo que ejemplifican los trabajos del Robotic Life Group, del MIT⁶, o las *Life Machines*, del mismo centro de investigación. La pauta básica ha consistido en imitar insectos, los cuáles son más simples (pero valga la complejidad de semejantes sistemas ‘simples’) que los animales, puesto que su relación con el mundo se limita la mayor parte de las veces a la percepción de señales, las cuales pueden considerarse como alarmas y activar un proceso reactivo hasta que la acción se equilibra nuevamente con el ambiente. Una mosca dirigiendo una escapada en un vuelo disuasorio ante una amenaza (un ser humano con un mata-moscas, por ejemplo) desarrolla una acción primaria de *alerta* o *miedo* ante el estímulo. Con ello tiene suficiente para su supervivencia. El desarrollo de emociones más complejas responde a la posibilidad de interaccionar de forma más rica tanto con el ambiente como con el resto de seres de la propia especie.

⁵ Accesible en versión en red: <http://www-ai.ijs.si/eliza/eliza.html>

⁶ <http://robotic.media.mit.edu/>

Una aproximación evolutiva a la conciencia y las emociones (Panksepp 2005) nos da la clave para el desarrollo de una IA *bottom-up* neoevolucionista que se refleja en la naturaleza para el desarrollo de máquinas afectivas.

2.2. Emociones en cuerpos o sistemas computacionales.

Tal y como hemos visto, el interés por el estudio de las emociones sintéticas en robots e interfaces reside en el desarrollo de la capacidad de socialización de tales instrumentos por parte de los usuarios humanos, quienes conciben estos agentes artificiales como ‘agentes con emociones’. Por ello, los elementos emocionales presentes en su funcionamiento deben ser satisfactorios del mismo modo que lo son las relaciones sociales humanas. Al fin y al cabo, tales máquinas son nuestros compañeros sociales (Reeves & Nass 1996).

Además, un entorno artificial emotivo diseñado correctamente activará empáticamente a su usuario, del modo según el cual los diseñadores hayan planeado. Recordemos que la empatía tiene una ubicación neuronal (Brave et al 2005) y que conforma la base del sistema de aprendizaje humano al permitir la interpretación y expresión similar de acciones.

2.2.1. Físicos o virtuales: INTERFACES.

En el ámbito de las interfaces, sobresale como precursora a la par que impulsora de elementos emocionales, Rosalind W. Picard, fundadora y directora del Affective Computing Research Group (MIT), esponsorizado por el consorcio industrial Things That Think Consortium. Al pasarse a un entorno virtual mediatizado por interfaces, una parte cada vez mayor de actividades humanas se realiza frente la pantalla de terminales diversos. El desarrollo de interfaces más intuitivos, fáciles de usar y que motiven positivamente a sus usuarios ha sido el campo de batalla de una cantidad ingente de programadores, ingenieros y diseñadores (Oatley 2004, Norman 2004). Las emociones han demostrado ser un elemento fundamental en el diseño de interfaces útiles, de manera que la interacción humano-máquina sea provechosa. Por lo argumentado anteriormente respecto de las proyecciones sociales y emotivas de los seres humanos hacia las máquinas, cabrá esperar que los intermediarios entre las máquinas y los humanos sean modelos adaptados a la naturaleza humana. Las formas antropomórficas de los robots domésticos son un ejemplo de lo comentado. Las emociones implementadas en tales sistemas (para su simulación ante el usuario o reconocimiento de las mismas para una adecuación de la actividad) son el miedo, la felicidad, la

culpa, la sorpresa y la cólera (Breazeal 2003). Existe una controversia acerca de cuáles son las emociones básicas o primarias (de forma muy básica son placer-dolor) que deben intentar implementarse en sistemas artificiales, y ello es debido al debate equivalente dentro del estudio de las emociones humanas (Ortony et al 1988).

2.2.2. Las emociones distribuidas.

Si bien existe un paradigma conceptual relativo a la cognición distribuida que pretende revolucionar y mejorar nuestros modelos de análisis de la cognición humana, estos adolecen del error clásico del pensamiento filosófico occidental: ignoran la función de las emociones, por lo menos en los textos principales y más sistemáticos de la disciplina.

En segundo lugar, los modelos que contemplan la existencia de emociones en procesos cognitivos lo hacen desde una perspectiva intrasubjetiva, en lugar de intersubjetiva, es decir, que analizan el valor cognitivo de las emociones para el individuo, obviando un hecho fundamental, a saber: que las emociones tienen verdadero sentido en su interacción social, lo que se ha denominado el *social sharing of emotion* (Finkenauer & Rimé 1998), (Rimé et al, 1991, 1998), (Christophe & Rimé 1997).

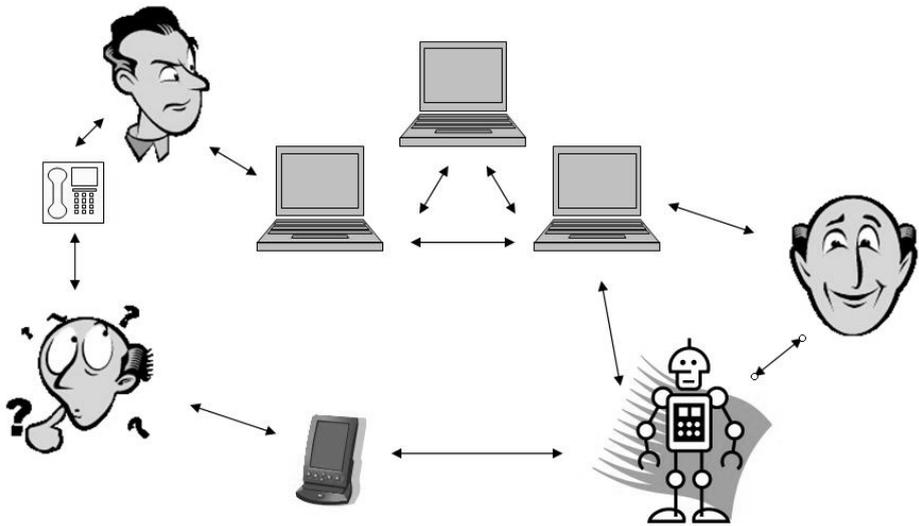
Desde esta perspectiva, las emociones experimentadas por una persona, considerando un amplio rango de niveles sociales, educativos y culturales, tienden a ser compartidas verbalmente con otras personas durante unos días por el 80-95% de la gente. Esto requiere una pregunta ¿por qué resulta necesario compartir las emociones? Tras cambios en el universo simbólico, los episodios emocionales son los verdaderos estimulantes de la búsqueda de respuestas para reestablecer el orden, por lo que forman una parte fundamental de los procesos cognitivos. Al mismo tiempo, el compartir emociones fortalece los lazos sociales. No por nada, se ha incluso diseñado una interface, AltarNation, de creación de espacios religiosos para comunidades virtuales (Hlubinka et al 2002).

Por ello, el diseño de sistemas afectivos artificiales, requiere de la consideración de entornos múltiples emocionales, los cuales participan activamente en un flujo emocional. Veamos en el siguiente punto una muestra de esta idea.

3. El ámbito de las emociones artificiales:

Tras lo expuesto, está claro que las emociones sintéticas están implementadas en diversos niveles de relaciones entre agentes participantes, considerando como agentes a estos elementos artificiales

con los que trabajan conjuntamente los seres humanos. Una red posible de relaciones entre agentes es la mostrada por la figura siguiente:



Abarca diversos ámbitos de relaciones, según un esquema adaptado de (McNeese 2003) y desarrollado por los autores:

	En un primer plano	En un segundo plano
Mediación Humano-Humano	Reconocimiento de estados emocionales (telemedicina,...) Dispositivos móviles inalámbricos Videoconferencias	Personalidades sintéticas 'Avatars' (AV)
Interacción Humano-Máquina	Interfaces gráficas Computadoras 'llevables' Captura y reacción emocional	Domótica Aprendizaje afectivo Reconocimiento de voz Inteligencia social
Interacción Máquina-Máquina	Robots sociables Coordinación sistemas afectivos	Webs semánticas

En este marco conceptual, los diversos agentes interaccionan emocionalmente en niveles distintos de complejidad emocional, aunque muestran la conexión de los diversos agentes en un modelo red de trabajo en red.

Este entorno emocional, nos muestra cómo los diversos intentos por generar entornos emocionales sintéticos desarrollan un papel importante y adecuado (por el momento) a las necesidades en las que se encuentran implementadas. El debate sobre las emociones, si bien se focaliza en tres ámbitos, a saber (1) la naturaleza de las emociones humanas, (2) las emociones sintéticas en el enfoque *top-down* (miméticas), (3) las emociones sintéticas *bottom-up* (neoevolucionistas), se reduce al fin y al cabo a la implementación de tales emociones en los modelos cognitivos que permiten la obtención de conocimiento.

4. Aplicaciones reales o posibles de las emociones sintéticas.

Llegados a este punto, nuestro análisis teórico sobre el papel de las emociones en los procesos racionales y la necesidad de crear modos de representarlas en los entornos artificiales presentes en la ciencia contemporánea, pasa por la descripción de cuatro posibles ámbitos de desarrollo:

4.1. Investigación científica.

Dentro del nuevo paradigma de la e-ciencia, definida ésta por (FECYT 2004) como: “las actividades científicas a muy gran escala que deberán desarrollarse cada vez más mediante colaboraciones globales distribuidas y accesibles a través de Internet. Este tipo de actividad científica cooperativa requerirá acceso a bancos de datos muy voluminosos y a recursos de computación de muy gran escala, además de prestaciones de visualización de alta calidad y otro tipo de herramientas”. Este marco metodológico y epistemológico se fundamenta en el uso de tecnologías computacionales para la obtención, visualización, análisis, cálculo, transmisión y debate de los datos científicos.

La implementación de diseños que contemplen las particularidades tanto cognitivas como emotivas (estando estas últimas incluidas en las primeras) de los seres humanos favorece el desarrollo de una actividad eficiente y correctamente adaptada a los entornos humanos. Cuanto más humanas son nuestras máquinas, mejor trabajamos con ellas, si bien es cierto que los diseños ‘a medias tintas’, como las máquinas no situadas en el mundo (aunque sí programadas para un bajo nivel de interacción con los usuarios: pongamos por caso, mediante mensajes de voz), resultan más desagradables que no próximas. Al fin y al cabo, de lo que

se trata es de diseñar entornos computacionales invisibles (Norman 1999), que no percibamos como algo ajeno a nosotros mismos y de los que no nos planteemos cuestiones relevantes.

Con respecto a los diversos niveles computacionales (sistemas expertos, robots...) partícipes en la actividad científica, remito a Remito al robot creado por los británicos Ross King y Stephen Oliver (de la University of Wales, UK) para la experimentación genética, que *formula hipótesis, diseña experimentos para evaluarlas, los lleva a cabo, ajusta sus teorías en función de los resultados y repite el ciclo*. Algo impresionante. Una vez creado y programado el robot es capaz de trabajar de forma autónoma y crear verdadero conocimiento. La máquina fue presentada en la revista *Nature* el 15 de enero de 2004, volumen 427, págs: 247-252, con el título “Functional genomic hypotheses generation and experimentation by a robot scientist”. El robot también contribuyó a demostrar los prejuicios cognitivos humanos al querer invertir en experimentos caros al considerar que sus resultados serán más prometedores. Se trata tan sólo de un ejemplo entre muchas otras iniciativas similares (Valdés-Pérez 1995, 1999) nos demuestra también como los ordenadores (fundamentalmente los sistemas expertos, tales como Logic Theorist, DENDRAL, CONGENH, KEKADA, AM, BACON MECHEM, ARROWSMITH, GRAFFITI o MPD/KINSHIP) están colaborando con los seres humanos en la creación de conocimiento científico válido. Recientemente, programas de pensamiento razonador automatizado han resuelto algunos problemas importantes de las matemáticas como el problema del álgebra de Robbins (McCune1997, explica el caso), los cuasigrupos (*quasigroups*, entre los años 1993-1996; los cuasigrupos son magmas – en álgebra abstracta, un magma es un tipo de estructura algebraica especialmente elemental - no vacíos donde la división es siempre posible), las álgebras booleanas ternarias (las TBA, en 1978; para un análisis más reciente remito a Wilkerson y otros 1990) y muchos otros (Loveland, 1999).

Y el Proyecto Genoma para la secuenciación del genoma humano, además de cuestiones de tipo ético (patentes, privacidad de los datos,...), tenía tres grandes retos: por un lado secuenciar rápida y eficientemente, por otro almacenar estos datos y, finalmente, encontrar sentido a la información obtenida y catalogada. En estos tres pasos tenemos la presencia completa y absoluta de sistemas computacionales o robotizados: ABI 3700 para la primera fase, GenBank y el algoritmo BLAST para la segunda y el sistema experto Otto para la tercera (identificaba y caracterizaba genes del genoma humano). Todas ellas no son tareas meramente mecánicas, algunas de ellas requieren el desarrollo de un

pensamiento artificial similar al del ser humano, aunque más rápido, fuerte y preciso. *Citius, altius, fortius*, pero para los sistemas computacionales, aunque contemplando los elementos emotivos (y cognitivos) básicos que permiten una correcta interacción de los usuarios con los instrumentos científicos ‘inteligentes’. Es por esta razón que desarrollamos ‘interfaces amigas’ o robots antropomorfos, es decir, androides. Pensemos que el software de visualización, *imaging* en el argot internacional, producirá el mismo efecto en la biología molecular de finales de los noventa, como hizo por otras disciplinas como las matemáticas o la medicina. Al aproximar el funcionamiento de las máquinas a las características de los seres humanos, éstas son incorporadas con mayor rapidez, facilidad y eficacia. Debemos pensar que el 60% de la información que recibe nuestro cerebro procede de la visión (Humphreys 2004: 112), y que, por lo tanto, la visualización de los datos que obtenemos del mundo condiciona nuestras teorías sobre el mismo. Proyectos internacionales como Physiome Project intentan desarrollar meta-lenguajes de programación que permitan trabajar de una forma más intuitiva a los programadores/usuarios y que sean de fácil conectividad entre sí. CellML o SBML son ejemplos de lo comentado.

4.2. Educación.

A partir de algunas características descritas sobre lo que son las emociones naturales y especialmente su papel en los procesos de motivación, es fácil argumentar como su inclusión en proyectos de e-learning no es algo baladí sino una mejora importante. Por un lado tenemos la transmisión de emociones desde el ordenador al usuario, para así facilitarle el proceso de aprendizaje, a la manera en que lo haría un profesor humano. Esta tarea que en principio podría parecer harto compleja, en realidad es más sencilla de lo que parece. Aquí resultan muy informativos los resultados de Reeves & Nass (1996). Como ya hemos visto anteriormente, la propuesta de Reeves y Nass es que consideramos a las computadoras por defecto socialmente como nuestros compañeros. Más específicamente, a través de una serie de tests, Reeves y Nass demuestran como tendemos a relacionarnos con las computadoras como si fueran personas o lugares reales. Así, por defecto, tendemos a asignarle una personalidad por defecto a un interface, aunque si racionalizamos sabemos que eso es absurdo. Reeves y Nass mostraron como somos educados frente a una computadora y así a la hora de valorar un programa de e-learning en un test vía computadora seremos más duros si hacemos el test en otra computadora que si lo hacemos en la misma computadora con la que hemos tomado el curso de e-learning.

Frente a la obsesión desde la industria de crear simuladores de personalidades cada vez más realistas, y con detalles, Reeves y Nass mostraron por el contrario que las representaciones esquemáticas facilitaban en realidad la suposición de encontrarnos ante una máquina con una personalidad y emociones. El desarrollo de simulaciones muy detalladas parece de hecho enturbiar este efecto de “intentional stance” a la Dennett que nos lleva casi automáticamente a suponer una intención detrás de un objeto inanimado cuando su output está lo suficientemente elaborado. (cfr. Dennett 1987). No es tanto los detalles del rostro sonriente lo que nos hará creer que una interfaz está alegre como que todo el comportamiento del sistema recuerde el output de una persona feliz. Otro ejemplo de esta actitud la tenemos en la reinterpretación emocional que los usuarios hacen de las acciones de sus perros AIBO descritas en el apartado 2.1.1.

Reconocer la emoción del estudiante es mucho más complejo, pero también más relevante en un proceso de aprendizaje. Picard (1997) desarrolla de forma inteligente el caso, mostrando lo frustrante que ha de resultar un profesor mecánico que una y otra vez nos dice “La respuesta no es correcta, vuélvalo a intentar” cuando y a es la decimotava vez que lo intentemos. Quizás la computadora no se canse, pero la persona sí, y se desmotiva. Es necesario por tanto, ofrecerle al sistema una forma de reconocer la emoción de la persona en cuestión, detectar el inicio de la frustración y decidir si pasar a otra pregunta, revisar lo que ya se había visto, intentar otro enfoque a la misma pregunta, etc. La forma más sencilla y eficaz de analizar las emociones del usuario es a través de una serie de dispositivos detectores de la conductividad eléctrica, el ritmo cardíaco y respiratorio, etc. Pero no hace falta argumentar mucho en relación al poco éxito que tendría un acercamiento así entre la mayoría del público. Parece más razonable suponer que sistemas basados en el tono de voz y la expresión facial, al ser mucho menos invasivos, resultarán más eficaces, aunque de todas formas, el hecho de que una máquina pueda detectar y registrar nuestras emociones también genera interesantes problemas de privacidad, así como una multiplicación del fenómeno rechazo –el efecto “yuck” que acostumbran a generar las nuevas tecnologías- al constatar como las máquinas podrían entrar incluso en la esfera privada de nuestros pensamientos.

Kismet es sin duda el paradigma del estado actual de este tipo de investigaciones. Probablemente, los resultados obtenidos a través del proyecto Kismet se irán introduciendo progresivamente en las nuevas generaciones de productos e-learning, especialmente en procesos donde la empatía es muy importante, como el aprendizaje de idiomas.

4.3. Interacciones humanos-máquinas y ocio.

Divertirse es evidentemente un conjunto de emociones, de manera que la relación entre emociones artificiales y ocio debería ser obvia. De nuevo, Picard (1997) contiene la teoría básica y algunos ejemplos prácticos de como podría organizarse una interacción así. El campo donde hay más investigación es la música. Aunque conocemos con suficiente detalle los mecanismos generales por los que la música genera emoción (cfr. Por ejemplo Casacuberta 2000) lo cierto es que no disponemos de ninguna teoría que permita, a partir de un ejemplo musical concreto establecer el tipo de emoción que un usuario medio sentirá. Este desconocimiento se debe, en buena parte, a nuestra falta de una teoría psico-neurológica detallada que explique de forma detallada como nuestro cerebro, a partir de datos meramente acústicos es capaz de construir frases musicales. Incluso fenómenos en apariencia tan obvios como el volumen o el color de un instrumento implican complejos ejercicios mentales de reconstrucción que todavía no sabemos reconstruir plenamente. El lector o lectora interesados en el tema puede consultar Cook (1999).

La aplicación paradigmática en este campo está descrita en Picard (1997): el usuario llega a casa, el sistema detecta su emoción y, en función de esta, empieza a emitir una serie de temas musicales que han sido catalogados emocionalmente para que le anime si está deprimido, le ayude a relajarse si está nervioso, le apacigüe si está enfadado, etc. Un sistema así tiene una clara contraparte en versión marketing: ser capaz de detectar qué temas estimulan a una audiencia, y de que forma, a partir de un detector artificial de emociones como el descrito en la sección anterior para establecer los momentos de frustración del usuario de un sistema de e-learning.

Hay también investigación en el otro lado de la ecuación: como conseguir que una computadora seleccione música a través de criterios emocionales o, si se nos permite la personificación, conseguir que una máquina “expresé” sus emociones mediante la música, de manera que su performance sea similar a la de un DJ humano. Un ejemplo de esta investigación es Casacuberta (2004), el desarrollo de un sistema así en un contexto de arte experimental, buscando un algoritmo generador de música que resultara bastante diferente a la forma en que un DJ humano selecciona temas y que al mismo tiempo ofreciera ciertas características de coherencia emocional, de manera que uno pudiera asignar, al estilo de Reeves y Nass, cierta emoción al programa y comentar “esta computadora ha de estar de muy mal humor ahora”.

4.4. Interacciones máquina-máquina.

En último lugar, deberíamos considerar un ámbito de creciente y necesario interés: el de las interacciones entre máquinas, tanto entre robots como entre ordenadores conectados entre sí. Cuando nos referimos a robots, estamos pensando en robots sociales (Brooks 2002) que trabajan desarrollando estrategias de acción de forma colectiva. Tom y Jerry (Brooks 1990) son ejemplos pioneros del desarrollo de robots sociales que interactúan entre sí para llevar a cabo tareas relativamente complejas. Para ello, deben desarrollar estrategias de cooperación. Desde el momento en que ciertos robots mantengan relaciones emocionales con seres humanos (Fellous & Arbib 2005) y, al mismo tiempo, estos robots tengan que cooperar con otros robots o máquinas, parece claro que lo más útil será permitir que las propias máquinas desarrollen intercambios de tipo emocional. No parece rentable o eficaz, desde un punto de vista práctico, crear interfaces de relación diferentes para ámbitos diferentes (humanos, por un lado, y máquinas, por el otro), puesto que ambos se interrelacionan de forma continua.

Al mismo tiempo, el desarrollo de webs semánticas eficientes tendrá que contemplar aspectos emocionales de los usuarios, además de una búsqueda semántica que supere el estado actual de la situación, algo que ya se está estudiando. En este caso, se requiere una arquitectura computacional de las emociones sintéticas (Scheutz & Sloman 2001), que permita a las máquinas tanto entender las emociones de los usuarios, como el desarrollo de una semiótica de las emociones que permita a las máquinas reaccionar emocionalmente de una forma coherente para sus usuarios.

Por último, el desarrollo de la e-Ciencia contemporánea que trabaja de forma creciente en *clusters* o *grids*, entornos multiagentes, requerirá una implementación de estos elementos afectivos para una eficiente utilización de tales herramientas y una relación positiva con los instrumentos con los que comparten tantas horas los investigadores. No resulta suficiente contar con la evidente capacidad humana para establecer relaciones de empatía y proyección emocional sobre las máquinas: éstas deben reaccionar, si no comprender, antes tales acciones de sus usuarios humanos o cooperadores artificiales en una acción epistémico compleja y computacionalmente distribuida.

5. Conclusiones y propuesta final.

El presente artículo ha revisado las teorías sobre el papel de las emociones en los procesos humanos sociales y racionales, mostrando la necesidad del desarrollo de emociones sintéticas de tipo diverso para una

mejor interacción entre humanos y máquinas. Lo que parece claro es que los seres humanos desarrollamos cada vez más un número importante de actividades bajo entornos computacionales, y que la inclusión de elementos emocionales sintéticos en los mismos mejoran la respuesta y efectividad de los usuarios.

Los procesos que implican reconocimiento de las emociones de las personas crean una serie de problemas relacionados con la privacidad que es necesario encarar desde el principio, así como cierto “efecto yuck” que tendrán muchos usuarios ante la idea de permitir que una máquina registre sus emociones en los momentos en que está disfrutando, triste, enfadado, etc. Sin embargo son estas aplicaciones las que un mayor potencial comercial tienen y podrían facilitar el desarrollo de inversiones en el campo de las emociones artificiales.

La cuestión crucial que nos planteamos es ¿son las emociones sintéticas emociones genuinas de tipo artificial o más bien constituyen simulaciones de emociones? La respuesta es ardua, puesto que en los seres humanos, las emociones están estrechamente relacionadas con la conciencia, e incluso hoy en día no tenemos claro si otros animales tienen emociones, en el sentido de ‘tener’ una emoción. Sin embargo, los seres dotados de un sistema nervioso complejo parecen experimentar diversos estados emocionales, por lo menos los dos más básicos: placer y dolor. Bajo esta premisa, nuestra propuesta, además de reclamar una mejora de los sistemas de emociones sintéticas, presentes tanto en interfaces o robots (para su relación humano-máquina o máquina-máquina), consideramos necesario el plantear la necesidad de crear estados emocionales básicos en entornos sintéticos. Crear entornos artificiales que tengan dolor o placer.

La mera simulación no es suficiente, aunque su trabajo es necesario, como demuestran los prometedores trabajos realizados por los autores (Vallverdú y Casacuberta 2008, 2009). Siguiendo las ideas de Dennett en Dennet (1986) la simulación de una tormenta no es una verdadera tormenta, después de todo, el ordenador no acaba mojado una vez se acaba la simulación. Igualmente, aunque hacer que una computadora simule las emociones puede ser útil para los proyectos descritos en la sección 4, a nivel teórico ello tiene muy poco que añadir a una modelización de las emociones desde la inteligencia artificial.

Es este un proyecto complejo que requiere la superación de diversos problemas técnicos y conceptuales, algunos de los cuales hemos pretendido resolver mediante el texto actual. Un entorno sintético emocional, permitiría disponer de un modelo controlable y fácilmente observable que nos facilitaría el estudio de la acción, partiendo de la

interacción emocional con el entorno, reproduciendo bajo una sucesiva estructura de capas de complejidad cognitiva los elementos que confluyen en la realización de acciones inteligentes. Nuestra reflexión persigue un modelo teórico completo sobre las emociones y su valor en los procesos cognitivos que se fundamente en un modelo computacional experimental.

Bibliografía

- Adamatzky, A., 2003, "Affectons: automata models of emotional interactions", *Applied Mathematics and Computation*, vol. 146, pp.579-594.
- Bates, J., 1994, "The role of emotions in believable agents", *Communications of the ACM*, vol. 37, nº 7, pp. 122-125.
- Batson, C.D. et al., 1997, "Empathy and attitudes: can feeling for a member of a stigmatized group improve feelings toward the group", *Journal of Personality and Social Psychology*, vol. 72, nº 1, pp. 105-118.
- Brave, S., Nass, C., Hutchinson, K., 2005, "Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent", *International Journal of Human-Computer Studies*, vol. 62, pp. 161-178.
- Breazeal, C., 1998, "Regulating human-robot interaction using 'emotions', 'drives' and facial expressions", presentado en el workshop *Autonomous Agents: Agents in Interaction – Acquiring Competence through Imitation*, Minneapolis/St Paul, Mayo.
- Breazeal, C., 2002, *Designing sociable robots*, MIT Press, Cambridge (MA).
- Breazeal, C., 2003, "Emotion and sociable humanoid robots", *Int. J. Human-Computer Studies*, vol. 59, 119-155.
- Brooks, R.A., 1990, "Elephants Don't Play Chess", *Robotics and Autonomous Systems*, vol. 6, pp. 3-15.
- Brooks, R.A., 1991, "Intelligence without representation", *Artificial Intelligence*, vol. 47, pp. 139-159.
- Brooks, R.A., 1994, "Building Brains for Bodies", *Autonomous Robots*, vol. 1, pp. 7-25.
- Brooks, R., 2002, *Flesh and Machines: How Robots Will Change Us*, Pantheon Books, NY.
- Casacuberta, D., 2000, *Qué es una emoción*. Crítica. Barcelona.
- Casacuberta, D., 2004, "DJ El niño: Expressing synthetic emotions with music", *Artificial Intelligence and Society*. Vol 18. pp. 187-191.
- Christophe, V., & Rimé, B., 1997, "Exposure to the social sharing of emotion: Emotional impact, listener responses and secondary social sharing", *European Journal of Social Psychology*, vol. 27, pp. 37-54.
- Colby, K.M., 1981, "Modeling a paranoid mind", *The Behavioral and Brain Sciences*, Vol. 4, pp. 515-560.
- Cook, P. R. (ed) 1999, *Music, Cognition and Computerized Sound: an Introduction to Psychoacoustics*, MIT Press, Cambridge (Mass).
- Damasio, A., 1995, *Descartes error*. MIT Press. Cambridge (Mass).
- Davis, R. & Lenat, D.B., 1980, *Knowledge-Based Systems in Artificial Intelligence*, McGraw-Hill, NY.
- DeLancey, C., 2001, *Passionate Engines: What Emotions Reveal About Mind and Artificial Intelligence*, Oxford University Press, Oxford.
- Dennet, D., 1986, *Brainstorms*. Haverster Press. Brighton.

- Dennet, D., 1987, *The Intentional Stance*, MIT Press, Cambridge (Mass).
- Ekman, B., 1973, "Cross-cultural studies of facial expressions" en Ekman (ed) *Darwin and Facial Expression: a Century of Research in Review*. Academic Press, New York.
- Elliot, C., 1992,, "The affective reasoner: a process model of emotions in a multi-agent system", Tesis doctoral Northwestern University, Mayo, *The Institute for Learning Sciences, Technical report #32*.
- FECYT, 2004, *Libro blanco. E-Ciencia en España*, Fundación Española para la Ciencia y la Tecnología, Madrid. [Disponible de forma gratuita como archivo PDF en <http://www.fecyt.es/documentos/e-Ciencia.pdf>].
- Fellous, Jean-Marc & Arbib, Michael A., 2005, *Who Needs Emotions? The Brain Meets the Robot*, Oxford University Press: Oxford.
- 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.
- Friedman, B., Kahn, P. & Hagman, J., 2003, "Digital Sociability: Hardware Companion? What Online AIBO Discussion Forums Reveal about the Human-Robotic Relationship", HI, *Proceedings of the conference on Human factors in computing systems*.
- Goleman, D. 1997. *Emotional intelligence*. Bantam Books. NY.
- Hey, T. & Trefethen, A.E., 2005, "Cyberinfrastructure for e-Science", *Science*, 308 (5723): 817-821.
- Hlubinka, M. et al, 2002, "AltarNation: Interface design for meditative communities", *CHI 2002 Extended Abstracts*, pp.612-613.
- Hollnagel, E, Woods, D.D., 1983, "Cognitive systems engineering: new wine in new bottles", *International Journal of Man-Machine Studies*, vol. 18, pp. 63-86.
- Horvitz, E.J., 1993, "Automated Reasoning for Biology and Medicine", en Fortuner E. (ed.), *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases and Computer Vision*, Johns Hopkins University Press, USA, pp. 1-40. (1993).
- Humphreys, Paul, 2004, *Extending Ourselves. Computational Science, Empiricism and Scientific Method*, Oxford University Press, Oxford.
- Hutchins, E., 1995, *Cognition in the Wild*, MIT Press, Cambridge (MA).
- Klein, J., Moon, Y., Picard, R.W., 2002, "This computer responds to user frustration: Theory, design and results", *Interacting with Computers*, vol. 14, pp. 119-140.
- Loveland, D.W., 1999, "Automated Deduction.Looking Ahead", *AI Magazine*, Volume 20 Number 1, pp. 77-98.
- Lyons, J., 1993, *Emoción*, Anthropos, Barcelona.
- McCune, W., 1997, "Solution of the Robbins Problem", *JAR*, 19(3), pp.263-276.
- McNeese, M.D., 2003, "New visions of human-computer interaction: making affect compute", *Int. J. Human-Computer Studies*, vol 59, pp. 33-53.
- Newell, A. & Simon, H.A., 1959, 'The Simulation of Human Thought', Report No. P-1734, The RAND Corporation, Santa Monica (CA).
- Newell, A., 1980, "Physical symbol systems", *Cognitive Science*, vol. 4, pp. 135-183.
- Norman, D.A., 1999, *The Invisible Computer*, MIT Press, Cambridge (MA).
- Norman, D.A., 2004, *Emotional design. Why we love (or hate) everyday things*, Basic Books, USA.
- Oatley, K., 2004, "The bug in the salad: the uses of emotion in computer interfaces", *Interacting with computers*, vol. 16, pp. 693-696.
- Oatley, K. & Jenkins, J. M., 1996, *Understanding Emotions*. Blackwell. London.

- Ortony, A., Clore, G.L., Collins, A., 1988, *The cognitive structure of emotions*, Cambridge University Press, Cambridge (MA).
- Panksepp, J., 2005, "Affective consciousness: Core emotional feelings in animals and humans", *Consciousness and Cognition*, vol. 14, pp. 30-80.
- Parasuraman, R., 1997, "Humans and automation: use, misuse, disuse, abuse", *Human Factors*, vol. 39, pp. 230-253.
- Picard, R., 1997, *Affective Computing*, MIT Press, Cambridge (MA). [www.nd.edu/~airolab/publications/AAAI104ScheutzM.pdf]
- Picard, R.W., & Klein, J., 2002, "The computer responds to user frustration: theory, design, and results", *Interacting with Computers*, vol. 14, n° 2, pp. 119-140.
- Plutchik, R., 1991, *The Emotions*, University Press of America, Lanham (MD).
- Reeves, B., & Nass, C.I., 1996, *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*, Cambridge University Press, Cambridge (MA).
- Rimé, B. et al, 1991, "Beyond the emotional event: six studies on the social sharing of emotions", *Cognition and emotions*, vol. 5, pp. 435-465.
- Rimé, B. et al, 1998, "Social sharing of emotion: new evidence and new questions", *European Review of Social Psychology*, vol. 9, pp. 145-189.
- Matthias Scheutz, M., 2004, "Useful Roles of Emotions in Artificial Agents: A Case Study from Artificial Life", *AAAI 2004*: 42-48
- Scheutz, M. & Sloman, A., 2001, "Affect and agent control: Experiments with simple affective states", en Zhong, N., Liu, J. Ohsuga, O., & Bradshaw, J. (eds.) *Intelligent agent technology: Research and development*, pp. 200-209, World Scientific Publisher, NJ.
- Schneiderman, Ben, 2008, "COMPUTER SCIENCE: Science 2.0", *Science*, Vol. 319. no.5868, pp. 1349 – 1350.
- Simon, H.A., 1997, "Artificial Intelligence: an empirical science", *Artificial Intelligence*, vol. 77, pp. 95-127.
- Sloman, A. & Croucher, M., 1981, "Why robots will have emotions", en *Proceedings 7th International Joint Conference on AI*, Morgan-Kaufman, USA.
- Solomon, R., 1999, "The Philosophy of Emotions", en Lewis & Haviland-Jones (eds), *Handbook of Emotions*, Guilford Press, NY, pp. 3-15.
- Thagard, P., 2002, "The passionate scientist: emotion in scientific cognition", en P. Carruthers, S. Stich, y M. Siegal (eds.), *The Cognitive Basis of Science*, Cambridge University Press, Cambridge.
- Tosa, N. & Nakatsu, R., 1996, "Life-like communication agent – emotion-sensing character 'MIC' and feeling session character 'MUSE'", en *Proceedings of the 1996 International Conference on Multimedia Computing and systems (Multimedia '96)*.
- Valdés-Pérez, Raúl E., 1995, "Machine discovery in chemistry: new results", *Artificial Intelligence*, vol. 74, pp. 191-201.
- Valdés-Pérez, Raúl E., 1999, "Principles of human-computer collaboration for knowledge discovery in science", *Artificial Intelligence*, vol. 107, pp. 335-246.
- Vallverdú, Jordi & Casacuberta, David (2008) "The Panic Room. On Synthetic Emotions" en Briggie, A., Waelbers, K. & Brey, P. (eds) *Current Issues in Computing and Philosophy*, 103-115, The Netherlands: IOS Press.
- Vallverdú, Jordi & Casacuberta, David (2009) "Modelling Hardwired Synthetic Emotions: TPR 2.0." en Vallverdú, J. & Casacuberta, D. (eds) (2009) *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence*, 103-115, USA: IGI Global. *En proceso de*

edición.

Weizenbaum, J., 1966, “A Computer program for the study of natural language communication between man and machines”, *Communications of the Association of Computing Machinery*, vol. 9, pp. 36-45.

Wilkerson, R., Murphy, D., Jenness, J., Pinet, B., Brazeal, D., 1990, “Ternary Boolean algebra unification”, *Applied Computing, Proceedings of the 1990 Symposium on*, Volume 5, Issue 6, pp.:119 - 121

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*. Routledge & Kegan Paul, London.

Recibido el 21 de julio de 2008; aceptado el 02 de abril de 2009.