



CURSO 2021-2022

APLICACIÓN DE MODELOS DE EMOCIONES A AGENTES

AGENTES INTELIGENTES

ANDREA GASCÓ PAU E IÑAKI DIEZ LAMBIES
UNIVERSITAT POLITÈCNICA DE VALÈNCIA
Escola Tècnica Superior de Enginyeria Informàtica



1 CONTENIDO

2	Introducción.....	2
3	Agentes	2
3.1	Agente BDI.....	2
3.2	Agente afectivo	3
3.3	Agentes normativos.....	3
4	Las emociones	4
4.1	Teorías de la emoción.....	4
4.1.1	Modelo OCC	5
4.2	Diferencias culturales.....	5
4.3	Emociones en la computación.....	6
4.3.1	Asignación de etiquetas emocionales	6
5	Agentes emocionales.....	8
5.1	Modelos de agentes emocionales	9
5.1.1	Normativos emocionales.....	9
5.1.2	Aprendizaje de conceptos emocionales.....	11
6	Conclusiones.....	15
7	Tabla de ilustraciones	16
8	Bibliografía	17

2 INTRODUCCIÓN

Las emociones juegan un importante papel en el ámbito comunicativo entre seres humanos, así como estos con su entorno. Es por ello por lo que podemos suponer que aquellos sistemas que sean capaces de trabajar con las emociones pueden mejorar su interacción haciéndola más natural y realista, tanto con su entorno como con otros humanos. (Taverner, Vivancios, & Botti, 2020)

Este trabajo trata de explorar una visión general de tanto la propia substracción del agente como su evolución hacia el acercamiento humano, al mundo de las emociones. Exploramos las diferentes alternativas que se han presentado a lo largo de los años, así como sus avances entre ellas.

3 AGENTES

Podemos encontrar una gran cantidad de definiciones de lo que es un agente. En primer lugar, los agentes software son aquellas aplicaciones informáticas que tienen capacidad de decidir cómo deben actuar para alcanzar sus objetivos. En segundo lugar, los agentes inteligentes son aquellos agentes software que pueden trabajar de una forma fiable en entornos cambiantes e impredecibles. (Botti, Carrascosa, & Julián)

Una de las características más importante que puede poseer un agente es su autonomía, y a raíz de esto, en (Wooldrige, 2009) hace una primera aproximación de lo que es un agente como un sistema informático que es capaz de actuar de manera autónoma en un entorno con el fin de lograr ciertos objetivos que le haya sido asignados. (Botti, Carrascosa, & Julián)

Actualmente, dentro de la teoría de agentes, podemos ver una definición más completa la cual sería que un agente es un sistema informático que se encuentra en algún entorno, el cual es percibido a través de entradas como sensores y a partir de tales percepciones decide, mediante alguna técnica de resolución de problemas, ejecutar determinadas acciones, de forma autónoma, que le permiten llegar a realizar con éxito sus objetivos y que pueden realizar modificaciones en el entorno en el que se encuentra. (Botti, Carrascosa, & Julián)

3.1 AGENTE BDI

Son agentes inteligentes que se basan en tres elementos: *beliefs* (creencias), *desires* (deseos) y *intentions* (intenciones).

Las creencias generalmente se definen como una convicción de la verdad de una proposición. Estas se pueden adquirir a través de la percepción, la contemplación o la comunicación. Es decir, es una proposición verdadera que debe corresponder a la realidad y derivarse de pruebas y argumentos. (Jiang, Vidal, & Huhns, 2007)

Por otro lado, los deseos apuntan a las opciones que están disponibles para el agente o los posibles cursos de acción de este. Se obtienen a través de una función de generación de opciones en base a sus creencias e intenciones actuales. (Jiang, Vidal, & Huhns, 2007)

En último lugar, las intenciones juegan un papel crucial ya que tienen como objetivo conducir la acción del agente. Representan el enfoque actual del agente, es decir, aquellos estados que se ha comprometido a intentar provocar y que se ven afectados por sus creencias actuales. (Jiang, Vidal, & Huhns, 2007)

3.2 AGENTE AFECTIVO

Sistema que es capaz de simular el comportamiento afectivo humano. Existen diversos enfoques con diferentes modelos de representación de las emociones como en: (Alfonso, Vivancos, & Botti, 2017), (Gebhard, 2005), (Marsella & Gratch, 2009), (Meuleman & Scherer, 2013) o (Reisenzein, y otros, 2013). (Taverner, Vivancios, & Botti, 2020)

3.3 AGENTES NORMATIVOS

Una norma es una regla que se debe seguir o a que se deben ajustar diferentes conductas, tareas o actividades. (Real Academia Española, 2022)

Las personas somos seres sociales para los que las emociones tienen un fuerte impacto, ya sea de manera positiva o negativa, afectando nuestras relaciones y toma de decisiones para una determinada acción, teniendo en cuenta cómo nos pueden afectar a nosotros mismos y a otras personas. En estas tomas de decisiones también toman un papel muy importante las normas o reglas y las consecuencias que estas pueden traernos. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

Por tanto, a estos efectos, podemos distinguir cuatro tipos de normas:

En primer lugar, las normas institucionales, las cuales se asemejan a las leyes establecidas por una autoridad. Dentro de los sistemas multiagente, estas normas están predefinidas, lo que quiere decir que se encuentran inicialmente en el conocimiento del propio agente o deben ser transmitidas. Aquí podemos hablar también del incumplimiento de una de estas normas, hecho que está asociado a un castigo o una sanción, como puede ser poner alguna restricción a las acciones que un agente puede realizar. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

En segundo lugar, las normas sociales, son aquellas que se crean a través de la sociedad. Estas normas van evolucionando, de abajo hacia arriba, debido a las interacciones o comportamientos entre los miembros de esta. A diferencia de las normas institucionales, estas normas no son creadas por ninguna autoridad, son, más bien, la representación de diferentes comportamientos que surgen a raíz que repetidas interacciones entre agentes. Para nosotros, si alguna persona se saltara alguna de estas normas como, por ejemplo, interrumpir a otra persona cuando habla, sería considerado como una persona irrespetuosa y los demás podrían hacer esas mismas acciones hacia su persona. Por tanto, el no seguir este tipo de normas, puede llevar a que el grado de confianza del resto de agente hacia ese agente concreto se vea mermado. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

En tercer lugar, las normas de interacción o normas de grupo son aquellas que se crean debido a acuerdos entre las partes. Son normas institucionales que se crean dentro de un subconjunto de agentes. Para nosotros serían como la equivalencia a la realización de un contrato o un acuerdo legal. Estas pueden ser creadas para un periodo de tiempo limitado, y se ve que acciones están permitidas y cuáles no. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

Por último, nos encontramos con las normas privadas, que son aquellas que se encuentran en la “mente” del agente. Son reglas internas autoimpuestas que aseguran la autonomía de este. Pueden ser creadas debido a la internalización de una norma social, por ejemplo, y pasan a ser como principios. Para nosotros, serían equivalentes a los comportamientos que una persona tiene consigo mismo. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

Cabe destacar, que las normas no son estáticas en el tiempo, sino que tienen un ciclo de vida que puede comenzar con la creación de dicha norma, siendo creada por una autoridad de la organización e integrándola en el agente a través del código. Otra forma de empezar a tener una norma es con la identificación de la norma por parte del propio agente en tiempo de

ejecución. Por tanto, si la norma ha sido creada por una autoridad, se debe de transmitir al agente para que este pueda identificarla e interiorizarla.

Después de esto, una vez identificada, la norma pasa a un estado de interiorización y aceptación. Una vez se ha interiorizado una norma puede ser reforzada debido a las interacciones del agente y posteriormente, si este reforzamiento se ve mermado, la norma perderá importancia y será con el tiempo olvidada. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

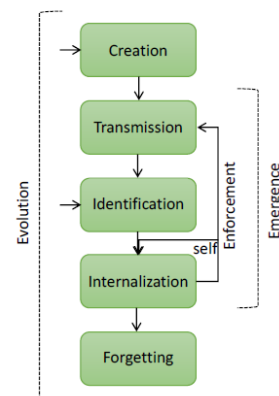


Ilustración 1 - Ciclo de vida de una norma (Argente, Botti, Perez-Garcia, & Del Val, 2020)

4 LAS EMOCIONES

Una emoción es una alteración del ánimo intensa y de tiempo limitado que se puede expresar a través de expresiones como la alegría, sorpresa, miedo, tristeza, ira o culpa, por ejemplo. Ligado a las emociones podemos hablar de las reacciones a diferentes estímulos que nos rodean y que se pueden manifestar a través de las emociones. (Ministerio de Sanidad, 2022)

Debido a la mejora del aprendizaje automático, existe una gran cantidad de modelos que puede reconocer emociones mediante imágenes voz, texto o incluso encefalogramas. Aun así, la mayoría de estos se enfocan en mejorar la precisión sin prestar atención a la forma de presentar ese conocimiento para su posible uso en sistemas como los agentes efectivos. (Taverner, Vivancios, & Botti, 2020)

4.1 TEORÍAS DE LA EMOCIÓN

Existen multitud de modelos, aunque uno de los más conocidos es *The Circumplex Model of Affect* propuesto en (Russell, 1980). Este relaciona las emociones respecto a valores de *Pleasure* (Placer) y *Arousal* (Excitación). Gracias a la experimentación se observó que las emociones siguen un patrón circular dentro de un espacio continuo basado en estas dos dimensiones.

El grado de intensidad de una emoción se puede también medir a través de este modelo. Según (Reisenzein R. , 1994) cuanto mayor es la intensidad de la emoción, mayor es el valor de las dimensiones *Pleasure* (Placer) y *Arousal* (Excitación). También se pudo deducir que, a valores muy bajos de estas dimensiones se puede suponer que no hay emoción.

En último lugar, se propuso agregar una tercera dimensión conocida como *Dominance* al espacio bidimensional anteriormente descrito. Este modelo fue conocido por sus siglas en inglés *PAD* y es uno de los más utilizados para diseñar agentes afectivos. Gracias a esta dimensión adicional se aumenta el espacio de representación permitiendo desambiguar algunas emociones. (Mehrabian, 1996)

Hoy en día no existe un consenso sobre cuál es la mejor manera de representar las emociones. Esto se debe a que se trata de un concepto muy complejo que depende de multitud de factores (personales, contextuales, culturales...). En cualquier caso, es necesario diseñar un modelo computacional que represente este conocimiento de la forma más flexible y precisa posible. (Taverner, Vivancios, & Botti, 2020)

Las teorías más relevantes son la teoría de valoración, la teoría dimensional y la teoría jerárquica que se diferencian en la forma en la que las emociones están representadas. (Taverner, Vivancios, & Botti, 2020)

La primera afirma que las emociones se generan a partir de las explicaciones e interpretaciones de las circunstancias de un individuo, es decir, cómo los individuos se relacionan con su entorno. Así, las emociones surgen de la evaluación de su entorno que impacta directamente en las creencias, objetivos y planes del agente. Uno de los modelos más relevantes de esta teoría es el modelo OCC. (Taverner, Vivancios, & Botti, 2020)

La segunda considera las emociones desde un punto de vista estructural, diferenciando sobre diversas dimensiones fundamentales como pueden ser la valentía y la excitación. Un estado afectivo concreto se representa como unas coordenadas dimensionales y son modelos fáciles de procesar e interpretar por sistemas informáticos. (Taverner, Vivancios, & Botti, 2020)

La última y tercera de estas teorías, considera un conjunto de emociones muy reducido con emociones básicas, que se consideran elementos básicos que permiten la construcción de emociones más complejas. (Taverner, Vivancios, & Botti, 2020)

4.1.1 Modelo OCC

Este modelo parte de la consideración de que existen tres grandes bloques de emociones donde se clasifican 22 emociones (Ortony, Clore y Collins, 1988) y se basa en la manera en la que los agentes pueden percibir el entorno. (Taverner, Vivancios, & Botti, 2020)

El primer bloque tiene que ver con las emociones relacionadas con las consecuencias de los eventos, como podría ser la alegría. El segundo se relaciona con las acciones de los agentes, como el orgullo. El tercer y último bloque está relacionado con los aspectos de los objetos o propiedades que se les atribuyen a estos, como puede ser el amor y el odio. (Taverner, Vivancios, & Botti, 2020)

4.2 DIFERENCIAS CULTURALES

Según diversos estudios, una misma etiqueta de emoción se puede interpretar de forma distinta dependiendo del idioma y la cultura del hablante. (Russel, Lwicka, & Niit, 1989) (Boiger, y otros, 2018) (Cordaro, y otros, 2018)

Por otro lado, según el constructivismo, las etiquetas que usamos para referirnos a las emociones dependen de la cultura y el lenguaje que se utilice, así pues, no existe una correspondencia directa para todas las etiquetas de emociones entre diferentes idiomas. (Hoemann, Xu, & Barrett, 2019)

J. A. Russel es uno de los teóricos constructivistas más citados. Presenta su teoría *Core-Affect* donde indica la importancia del lenguaje y la cultura a la hora de catalogar las diferentes emociones (Russel, Lwicka, & Niit, 1989)

En el estudio (Jackson, y otros, 2019) se analizaron unos 2500 idiomas para determinar el nivel de similitud de 24 etiquetas de emociones en diferentes culturas. Esto se relacionó con seis dimensiones asociadas a las emociones: *Pleasure, Arousal, Dominance, Certainty, Approach-avoidance, and Sociality*. De entre las anteriormente descritas, se detectó que las dos primeras son las más importantes para determinar la semántica de las emociones entre los diferentes humanos, es decir, comunes a todos los humanos. (Taverner, Vivancios, & Botti, 2020)

Si seguimos un modelo de representación basado en el descrito por (Russell, 1980) debemos de tener en cuenta estudios como (Cordaro, y otros, 2018) y (Lim, 2016). Según estos estudios, diferentes culturas pueden atribuir diferentes valores de *Pleasure* (Placer) y *Arousal* (Excitación) para la misma etiqueta de emoción. (Taverner, Vivancios, & Botti, 2020)

4.3 EMOCIONES EN LA COMPUTACIÓN

La mayor parte de las propuestas en el área de la computación afectiva utilizan el modelo emocional descrito en (Ekman, 1992) para definir las emociones, aunque muchos otros utilizan una representación dimensional de las emociones, como en el caso de (Bakhtiyari & Husain, 2014). Además, aquellos que utilizan la representación dimensional asumen que una emoción básica puede ser universalmente representada por un simple punto. (Taverner, Vivancios, & Botti, 2020)

Por otro lado, parece haber una tendencia a definir los modelos afectivos utilizando lógica difusa. De esta forma se puede aproximar la representación artificial de las emociones de forma que los humanos las expresamos.

Esto se puede apreciar en trabajos como (Jain & Asawa, 2015) donde se utiliza para manejar la incertidumbre en la evaluación de cinco de las emociones básicas de Ekman. También en (Kazemzadeh, 2010) propone una metodología para traducir etiquetas emocionales del español al inglés utilizando esta lógica, obteniendo las respectivas funciones de pertenencia difusa para cada emoción. (Taverner, Vivancios, & Botti, 2020)

A partir de *The Circumplex Model of Affect*, trabajos como (Ivonin, Chang, Chen, & Rauterberg, 2012) proponen un modelo de representación basado en vectores de dimensiones *Pleasure* (Placer) y *Arousal* (Excitación). Aun así, estos desplazan el origen de coordenadas desdibujando el concepto de *Sleepiness/Deactivation*, lo cual provoca una descontextualización de emociones que dependen de esta dimensión. Esto se puede solucionar mediante el uso de estadísticas circulares. (Taverner, Vivancios, & Botti, 2020)

Es común en la informática afectiva que el uso de emociones suela simplificarse mediante el uso de enfoques categóricos que reducen las emociones a etiquetas simples (como las seis propuestas por (Ekman, 1992)). De esta forma se pierde parte de información como la intensidad o la proximidad a otra emoción. En un espacio multidimensional donde las emociones puedan moverse a través de variables continuas darían una mayor cantidad de información que si solo se usaran etiquetas. (Taverner, Vivancios, & Botti, 2020)

Así pues, un modelo más adecuado sería definir las emociones como un área en el espacio dimensional de la emoción. De esta forma podrías representar mejor su complejidad y las similitudes entre múltiples emociones. (Taverner, Vivancios, & Botti, 2020)

4.3.1 Asignación de etiquetas emocionales

En (Taverner, Vivancios, & Botti, 2020), uno de los artículos de referencia para este trabajo, se realiza un trabajo de análisis previo, así como dos experimentos. El primero de estos trata de encontrar las mejores traducciones del inglés al español para las nomenclaturas de las categorías afectivas *Pleasure* y *Arousal*, así como de sus contrarias *Misery* y *Sleepiness*, respectivamente. En este llegaron a la conclusión de que los conceptos tenían las siguientes mejores traducciones (de mejor a peor dentro de las aceptables):

- *Pleasure*: placer, satisfacción o disfrute.
- *Arousal*: alertado.
- *Misery*: miseria o tristeza.
- *Sleepiness*: somnolencia o adormecimiento.

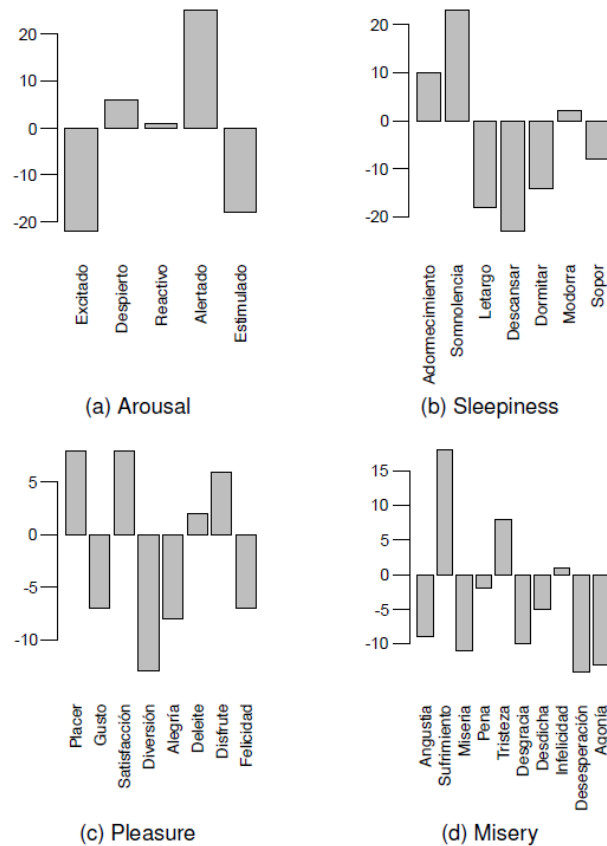


Ilustración 2 - Grado de aceptación de los candidatos para cada categoría afectiva (Taverner, Vivancios, & Botti, 2020)

En el segundo experimento trataron de asignar diferentes áreas a cada emoción a las diferentes emociones a partir de los niveles de las traducciones oportunas de *Pleasure/Misery* y *Arousal/Sleepiness* que obtuvieron del primer experimento. A partir de esto trataron de asignar a las diferentes emociones un área en el espacio dimensional a partir del análisis estadístico de los resultados.

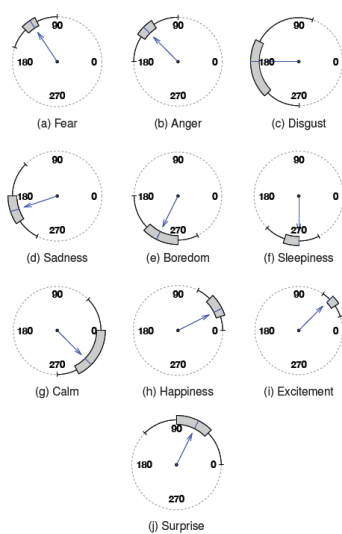


Ilustración 3 - Gráfica circular con vertical "Pleasure" y horizontal "Arousal" (Taverner, Vivancios, & Botti, 2020)

A partir de estos resultados confeccionan un modelo basado en lógica difusa que tiene en cuenta estos resultados donde, para clasificar una emoción, se tiene en cuenta su posición en espacio bidimensional, así como también la intensidad de esta.

Esto es así ya que, entre otras cosas, si una emoción tiene muy poca intensidad, podemos considerar que en realidad no hay emoción. Esto se consigue a través de una representación por vectores donde su módulo indica la intensidad y su dirección las posibles emociones que se están representando. Para discernir entre una emoción u otra se utiliza un modelo basado en lógica difusa. (Taverner, Vivancios, & Botti, 2020)

Entre las diferentes conclusiones de (Taverner, Vivancios, & Botti, 2020), cabe destacar la que dicta que, dependiendo de la cultura y lenguaje de la persona, esta puede asignar un espacio diferente en la representación dimensional de las emociones. Lo cual es de gran importancia para el desarrollo de agentes emocionales. Estos deberían tener en cuenta, tanto para

reconocer como para generar, la cultura o lenguaje con el que están trabajando para así realizar una correcta interpretación de las emociones.

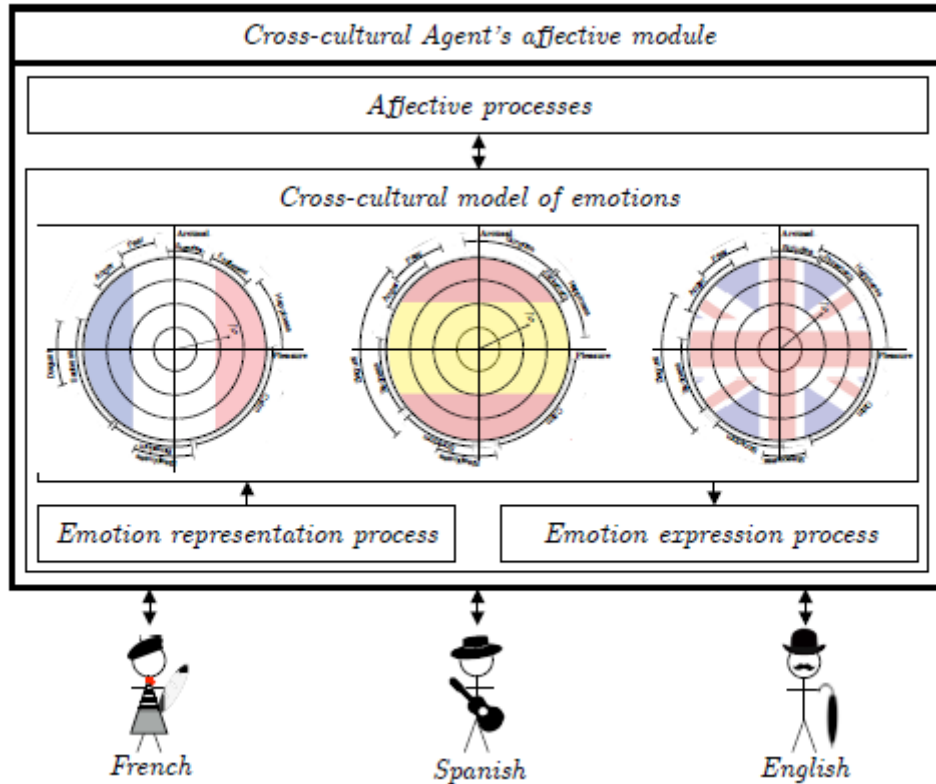


Ilustración 4 - Ejemplo de agente afectivo multicultural (Taverner, Vivancios, & Botti, 2020)

5 AGENTES EMOCIONALES

Un agente emocional añade a sus componentes BDI las emociones como elemento adicional. Estas actualizan teniendo en cuenta tanto las emociones primarias como las secundarias, tal como se describen en (Damasio, 1994). Es por ello por lo que se debe revisar el ciclo de vida del agente, incluyendo a sus funciones las emociones como componente adicional. (Jiang, Vidal, & Huhns, 2007)

Las emociones primarias son aquellas que sentimos primero como respuesta a una situación. Ejemplos de ello son el miedo al estar amenazados o la tristeza frente al concepto de la muerte. Preceden por tanto al pensamiento consciente y, por ello, son de naturaleza reactivas. Por todo esto, en el agente emocional encontraremos que se calcularán estas emociones como uno de los primeros pasos y sin tener en cuenta las creencias que no sean la percepción o la comunicación de ese bucle. (Jiang, Vidal, & Huhns, 2007)

EBDI-MAIN-LOOP

```

1   $E \leftarrow E_0;$             $\triangleright E_0$  are initial emotions
2   $B \leftarrow B_0;$             $\triangleright B_0$  are initial beliefs
3   $I \leftarrow I_0;$             $\triangleright I_0$  are initial intentions
4  while true
5      do  $B_p \leftarrow brf\_see(Env);$ 
6          $B_m \leftarrow brf\_msg(Cont);$ 
7          $E \leftarrow euf1(E, I, B_p \cup B_m);$ 
8          $B \leftarrow brf\_in(E, I, B \cup B_p \cup B_m);$ 
9          $D \leftarrow options(B, I);$ 
10         $I \leftarrow filter(E, B, D, I);$ 
11         $E' \leftarrow E$ 
12         $E \leftarrow euf2(E, I, B);$ 
13        if time permits and  $E \neq E'$ 
14            then  $B \leftarrow brf\_in(E, I, B);$ 
15                 $D \leftarrow options(B, I);$ 
16                 $I \leftarrow filter(E, B, D, I);$ 
17         $\pi \leftarrow plan(I, Ac);$ 
18        execute( $\pi$ )

```

Ilustración 5 - Seudocódigo de un agente EBDI (Jiang, Vidal, & Huhns, 2007)

Por otro lado, las emociones secundarias aparecen después de las primarias. Estas pueden ser causadas por las primeras o en consecuencia a complejas cadenas de razonamiento. Ejemplo de este tipo de emociones puede ser la ira que se genera posterior al miedo de un enfrentamiento. Las emociones secundarias pueden remplazar a las primarias en la toma de decisiones si el tiempo lo permite. (Jiang, Vidal, & Huhns, 2007)

5.1 MODELOS DE AGENTES EMOCIONALES

En este apartado desarrollaremos dos modelos de agentes emocionales que por su naturaleza significan un avance significativo en el estudio de estos conceptos.

5.1.1 Normativos emocionales

En apartados anteriores hemos hablado de los tipos de normas existentes y como se relacionan brevemente con los sistemas multiagente.

Dentro de los sistemas multiagente, las normas se han tenido en cuenta principalmente para solucionar algunos problemas con la coordinación y la cooperación entre agentes. Estas permiten ajustar los comportamientos e interacciones del agente con los demás, y definir las obligaciones, prohibiciones y demás acciones que deben de realizar. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

Podemos ver ahora, por tanto, como es un agente concreto que sigue este tipo de normas para relacionarse con él mismo, con el ambiente y con otros agentes de su entorno según a las normas que se rijan.

En primer lugar, nos encontramos con el agente “licito”, que es aquel que principalmente sigue las normas institucionales. En segundo lugar, el agente social es aquel que da preferencia a las normas sociales. Y, por último, el agente privado es aquel que juzga por él mismo y sigue las normas privadas. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

A pesar de que se ha analizado mucho en el análisis tanto de las emociones como de las normas, esto ha sido un proceso bastante independiente el uno del otro y, por tanto, la relación entre las emociones y las normas prácticamente no se ha considerado en los sistemas multiagente y, por tanto, la mayoría de los agentes normativos no consideran las emociones. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

Podemos encontrar diferentes trabajos para incorporar diversas teorías emocionales, como puede ser el modelo OCC, juntos con arquitecturas de agente y la representación de normas para definir así, una nueva arquitectura emocional que incluya tanto normas como agentes. Algunas de estas son, por ejemplo, TABASCO_{JAM}, o OP-NRD-E.

La primera mencionada, usa una representación de las normas *if-then-else*, donde se tiene en cuenta la consecuencia de una acción, lo honorable de un agente o el atractivo de un objeto. Donde la regla determina la posible generación de un estado emocional en consecuencia. Para esto, se han usado simplemente cuatro emociones principales (ira, alegría, resentimiento y piedad) y las normas se integran directamente en los agentes. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

La segunda mencionada, utiliza el modelo OCC considerando solo las emociones de alegría, orgullo, angustia y vergüenza para representar las emociones positivas y negativas. Además, usa un marco normativo donde las normas son vistas como obligaciones para la realización de una acción determinada a cumplir dentro de un plazo de tiempo restringido. En esta, las emociones se dan cuando ocurren los acontecimientos y el agente necesita usar ese recurso para conseguir su acción determinada. Podría ser, por ejemplo, cuando el agente alcanza su objetivo este puede demostrar orgullo, y en contraposición, cuando este no lo consigue podría

sentir angustia. Esto significa que los cambios en los eventos de los acontecimientos provocan emociones positivas y negativas en el agente. Si la emoción que ha representado es de ámbito negativo, el agente reevalúa sus planes, lo que significa que las emociones hacen que los agente consideren acciones alternativas para conseguir su objetivo. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

Se pueden describir cuatro tipos de relaciones entre las emociones y las normas.

En primer lugar, las emociones son tenidas en cuenta en el proceso de razonamiento normativo. Esto significa que el cumplimiento o no de una norma no solo debe basarse en las decisiones racionales, sino que se deben tener en cuenta las emociones para poder obtener una solución más realista al problema. Se deben considerar la utilidad esperada de esta decisión, la coherencia de la decisión tomada respecto a la cognición del agente y las emociones del agente respecto a sus acciones. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

En segundo lugar, la violación o el cumplimiento de una norma genera emociones en el propio agente que realiza la acción o en otro agente observador de dicha acción. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

En tercer lugar, las emociones son utilizadas como una forma de conseguir cumplir las normas sociales. Esto significa que si, por ejemplo, un agente viola una norma social y este desarrolla una emoción negativa, este hecho servirá de mecanismo para reforzar las normas sociales. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

Por último, las emociones ayudan a interiorizar las normas, lo que hace que determinadas acciones o conductas aceptadas influyan en la internalización y el cumplimiento de alguna norma en el propio agente. Esto puede acabar siendo visto como la aceptación de una norma privada del propio agente. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

A raíz de esto, (Argente, Botti, Perez-Garcia, & Del Val, 2020) proponen un modelo extensión de la arquitectura BDI para un Agente Normativo Emocional. Esta propuesta incluye un componente normativo, basado en el ciclo de vida mencionado con anterioridad y que sigue un proceso de identificación, internalización, razonamiento normativo y por último, el proceso del cumplimiento de la norma; un componente emocional compuesto por diferentes funciones, la valorativa, la generadora de emociones, generadora de afecto y la función de afrontamiento, y un componente BDI que contiene los elementos y procesos típicos de la arquitectura BDI conectados con los dos anteriores a través de diferentes funciones. La de percepción de revisión de creencias, la de mensaje de revisión de creencia, la generadora de opciones, la de deliberación y la selección de acciones. Con todo esto, se incluye el estado de ánimo esperado como la emoción provocada y, además, refleja las cuatro relaciones entre normas y emociones mencionadas anteriormente. (Argente, Botti, Perez-Garcia, & Del Val, 2020)

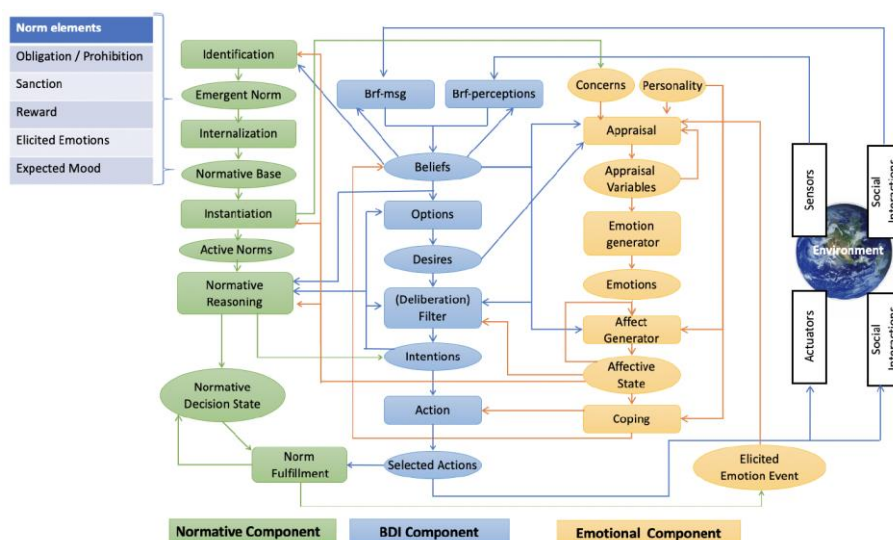


Ilustración 6 - Propuesta de arquitectura abstracta NEA (Argente, Botti, Perez-Garcia, & Del Val, 2020)

5.1.2 Aprendizaje de conceptos emocionales

En (Smith, Parr, & Friston, Simulating emotions: An active inference model of emotional state inference and emotion concept learning, 2019) se presenta como una investigación con el objetivo de construir un modelo computacional con la capacidad de aprender, reconocer y comprender las propias emociones.

Según la literatura sobre la habilidad de la conceptualización de las emociones, esta relaciona niveles más bajos de conceptualización a varios trastornos psiquiátricos, así como a una peor salud física. (Levine, Marziali, & Hood, 1997) (Berthoz, Ouhayoun, & Parage) (Moeller, y otros, 2014) (Lackner, 2005) Por otro lado, los niveles más altos se asocian con una gran variedad de rasgos y habilidades adaptativos relacionados con las emociones. (Ciarrochi, Caputi, & Mayer, 2003) (Baslet, Termini, & Herbener, 2009)

Es por esto por lo que ciertas modalidades psicoterapéuticas tienen como objetivo mejorar la comprensión de las emociones como parte central de la psicoeducación en psicoterapia. Aquí se puede apreciar, por tanto, lo importante de este estudio. (Hayes & Smith, 2005)

Generalmente se afirma que los conceptos de emoción deben adquirirse a través de la experiencia. Las teorías de emociones básicas sostienen que, emociones como la tristeza y el miedo, tienen circuitos neuronales distintos, pero no niegan que deben aprenderse. (Panksepp & Biven, 2012)

Por otro lado, puntos de vista constructivistas sostienen que las emociones no tienen una relación 1 a 1 con diferentes circuitos neuronales y que la adquisición del concepto de emoción es necesario para la experiencia emocional. (Barrett, 2017)

Con esta información, en (Smith, Killgore, & Lane, The structure of emotional experience and its relation to trait emotional awareness: a theoretical review, 2018) proponen un modelo de tres procesos donde se realiza un despliegue de conceptos emocionales para comprender las propias respuestas afectivas. Este modelo distingue los siguientes tres procesos:

1. Generación de respuestas afectivas

Los estados mentales y cognitivos se modulan para conseguir una respuesta en base a una evaluación de la importancia del estímulo y el logro de objetivos del individuo.

2. Representación de respuestas afectivas

Proceso percibe sus sensaciones y conceptualiza estas como una emoción particular teniendo en cuenta todas las demás fuentes disponibles de información (estímulo, pensamientos, creencias...).

3. Acceso consciente

Informes explícitos de su situación emocional actual.

Es aquí donde cabe presentar el concepto de inferencia afectiva. Partiendo del supuesto que el cerebro es una máquina de inferencia afectiva, se postula que el cerebro encarna un modelo interno del mundo generativo (simula datos sensoriales que debería recibir si su modelo fuera correcto). A partir de estas simulaciones se puede comprobar con la actividad sensorial actual y las desviaciones frente a la predicción se pueden utilizar para actualizar el modelo. (Smith, Parr, & Friston, *Simulating emotions: An active inference model of emotional state inference and emotion concept learning*, 2019)

A corto plazo esto se puede ver como la simple percepción mientras que a largo plazo se corresponde con el aprendizaje (actualizar las expectativas en pruebas posteriores). Esto afirma que el modelo generativo sigue siendo un modelo preciso del mundo. (Conant & Ashbey, 1970)

Entonces, la acción puede entenderse como un proceso mediante el cual el cerebro busca evidencia para sí mismo. Esto es así ya que hay un sentido profundo en el que el cerebro busca continuamente observaciones que respalden o sean internamente consistentes. (Hohwy, 2016)

Así pues, la inferencia afectiva propone que el cerebro calcule una cantidad estadística llamada *energía libre*. La minimización de esta es equivalente a la maximización de la evidencia del modelo. La toma de decisiones de un agente puede, por tanto, evaluar la energía esperada y aprender que políticas alternativas seleccionar en trayectorias futuras (calculando los estados futuros que se esperarían con cada política). (Smith, Parr, & Friston, *Simulating emotions: An active inference model of emotional state inference and emotion concept learning*, 2019)

La toma de decisiones será óptima si opera infiriendo la política que minimiza la energía libre esperada. Esta se puede descomponer en incertidumbre y preferencias previas. De esta forma, se seleccionarán primero políticas que minimizan la incertidumbre y luego aquellas que maximicen sus preferencias previas. (Smith, Parr, & Friston, *Simulating emotions: An active inference model of emotional state inference and emotion concept learning*, 2019)

A partir de este análisis, en (Smith, Parr, & Friston, *Simulating emotions: An active inference model of emotional state inference and emotion concept learning*, 2019) se detalla un modelo generativo formulado como un proceso de decisión de Markov parcialmente observable. Este modelo especifica las políticas permitidas, los estados ocultos del mundo y los resultados observables, así como una serie de matrices que definen las propiedades probabilísticas entre estos.

En la matriz *A* se especifica qué resultados se generan por cada combinación de estados ocultos (mapeo de que se observe un conjunto particular de resultados). En *B* se codifican las transiciones de estado, especificando de que un estado evoluciones a otro con el tiempo (algunas de estas son controladas por el agente dependiendo de su política). Por otro lado, en *D* se codifican las expectativas previas sobre el estado oculto inicial del mundo mientras que en *E* se hacen sobre qué políticas se elegirán (comportamientos habituales tienen mayor expectativa previa). Por último, en *C* se codifican las preferencias previas sobre los resultados. (Smith, Parr, & Friston, *Simulating emotions: An active inference model of emotional state inference and emotion concept learning*, 2019)

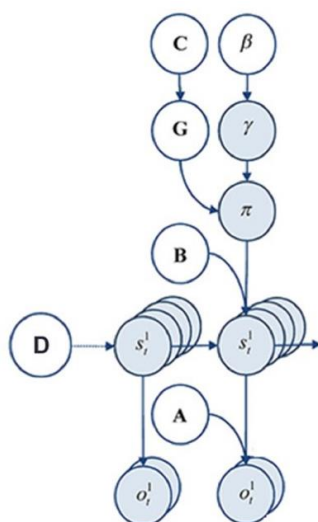


Ilustración 7 - Proceso de decisión de Markov de inferencia activa (Smith, Parr, & Friston, *Simulating emotions: An active inference model of emotional state inference and emotion concept learning*, 2019)

Por otro lado, para predecir los errores de predicción se calcula a través de un promedio la energía libre esperada de cada política y, a partir de esto, se selecciona la de menor error de predicción. (Smith, Parr, & Friston, *Simulating emotions: An active inference model of emotional state inference and emotion concept learning*, 2019)

Esto lleva a los investigadores a preguntarse el papel del aprendizaje en todo esto. Basándose en el estudio (Brown, Zhao, & Leung, 2009), se dan cuenta que la fuerza de la sinapsis podría codificar la cantidad de evidencia que proporciona una observación determinada (entrada de la matriz A), otra podría entonces también codificar la probabilidad de un estado en un momento posterior (entrada de la matriz B). (Smith, Parr, & Friston, *Simulating emotions: An active inference model of emotional state inference and emotion concept learning*, 2019)

Un experimento muy interesante que realiza el estudio principal de este apartado es el relacionado con el aprendizaje en base a la experiencia previa. En un punto de su estudio se preguntan si e pueden aprender los conceptos de emociones en la infancia a partir de investigar el modelo que han presentado.

Así pues, dan por supuesto que su modelo debería empezar sin creencias previas sobre como las emociones estructuran su experiencia (matriz A con un mapeo totalmente impreciso). Después de realizar 200 conjuntos de observaciones el modelo podía alcanzar con éxito el 100% de precisión y las transiciones del proceso generativo eran moderadamente altas. Lo cual acaba por probar el principio de que el modelo es bastante consistente de experiencias de su *infancia*. (Smith, Parr, & Friston, *Simulating emotions: An active inference model of emotional state inference and emotion concept learning*, 2019)

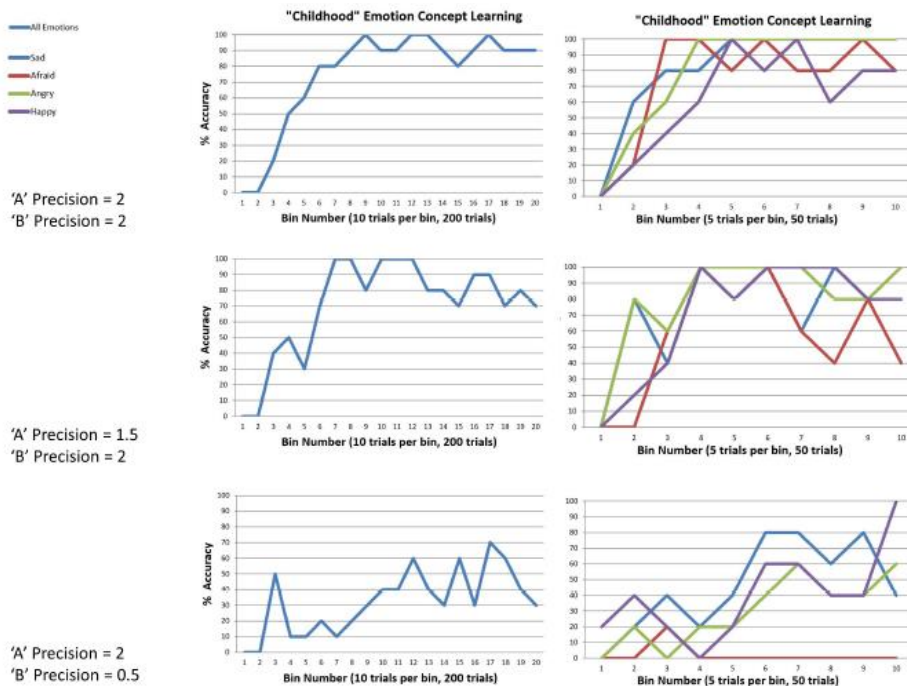


Ilustración 8 - Proceso de aprendizaje de la simulación "infancia" (Smith, Parr, & Friston, *Simulating emotions: An active inference model of emotional state inference and emotion concept learning*, 2019)

Ahora prosiguen a preguntarse si se podría realizar lo mismo en la edad adulta. Comienzan con una matriz A totalmente imprecisa, de nuevo, pero con la precisión de las matrices A y B a los niveles que en la infancia fueron exitosos. Después de 150 ensayos observamos que la precisión era alta al final de estos y, a partir de estos, los 200 siguientes se aprendieron las emociones con éxito. Cabe destacar que el aprendizaje de algunas emociones interfiere con otras (caso de *happy* con *sad*). (Smith, Parr, & Friston, *Simulating emotions: An active inference model of emotional state inference and emotion concept learning*, 2019)

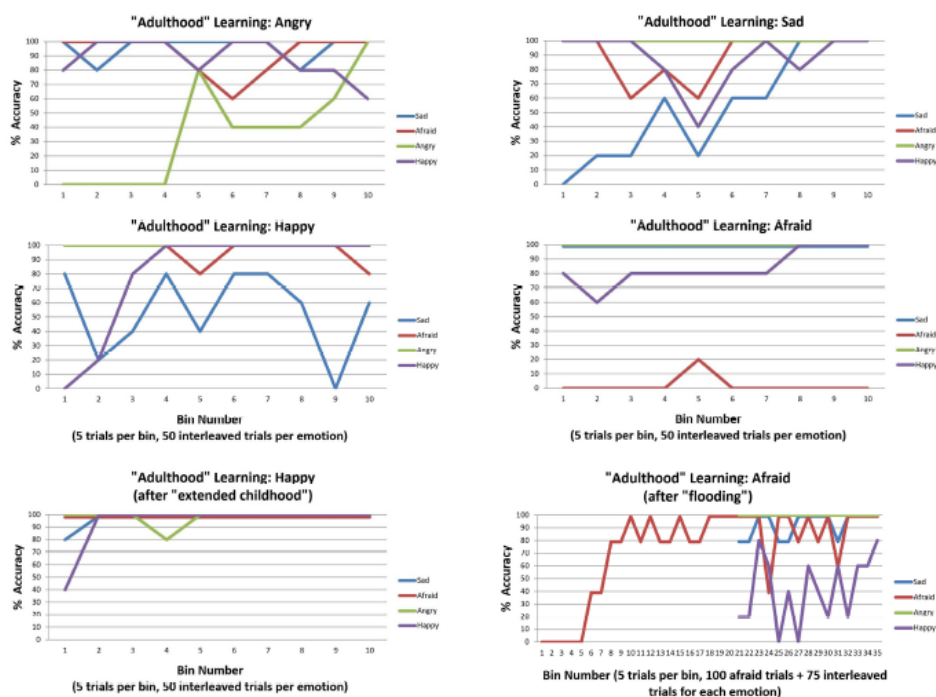


Ilustración 9 - Proceso de aprendizaje de la simulación "adulto" (Smith, Parr, & Friston, *Simulating emotions: An active inference model of emotional state inference and emotion concept learning*, 2019)

Modelos como este nos dan una visión muy interesante de como el estudio de agentes emocionales puede tener una gran relación con la psicología y, en concreto, con la psicoterapéutica.

6 CONCLUSIONES

Como hemos podido ver, el campo de los agentes emocionales es amplio y variado, así como con mucho recorrido por delante. Es apasionante como la psicología juega un papel crucial a la hora de ayudar al modelado de los agentes. Aplicando diferentes modelos emocionales se consiguen resultados cambiantes y, dependiendo de la mezcla de estos que se utilicen, incluso con mayor precisión.

La evolución de este tipo de modelados es clave en el futuro entendimiento computador-humano, cuanto más acerquemos el modelo del pensamiento humano al computador, más real será la interacción y, seguramente, más productiva y comunicativa. El campo de los agentes emocionales presenta, en nuestra humilde opinión, una aproximación más que acertada ya que, lo que principalmente diferencia un computador de un humano, es en su capacidad afectiva.

Este trabajo nos ha permitido conocer más en profundidad el campo de los agentes inteligentes y, en concreto, permitirnos conocer aquellos dedicados a simular las emociones. Es un campo apasionante e inabarcable por la cantidad de trabajo hecho y por el que queda por hacer.

7 TABLA DE ILUSTRACIONES

<i>Ilustración 1 - Ciclo de vida de una norma (Argente, Botti, Perez-Garcia, & Del Val, 2020)</i>	4
<i>Ilustración 2 - Grado de aceptación de los candidatos para cada categoría afectiva (Taverner, Vivancios, & Botti, 2020)</i>	7
<i>Ilustración 3 - Gráfica circular con vertical "Pleasure" y horizontal "Arousal" (Taverner, Vivancios, & Botti, 2020)</i>	7
<i>Ilustración 4 - Ejemplo de agente afectivo multicultural (Taverner, Vivancios, & Botti, 2020)</i>	8
<i>Ilustración 5 - Seudocódigo de un agente EBDI (Jiang, Vidal, & Huhns, 2007)</i>	8
<i>Ilustración 6 - Propuesta de arquitectura abstracta NEA (Argente, Botti, Perez-Garcia, & Del Val, 2020)</i>	11
<i>Ilustración 7 - Proceso de decisión de Markov de inferencia activa (Smith, Parr, & Friston, Simulating emotions: An active inference model of emotional state inference and emotion concept learning, 2019)</i>	13
<i>Ilustración 8 - Proceso de aprendizaje de la simulación "infancia" (Smith, Parr, & Friston, Simulating emotions: An active inference model of emotional state inference and emotion concept learning, 2019)</i>	13
<i>Ilustración 9 - Proceso de aprendizaje de la simulación "adulto" (Smith, Parr, & Friston, Simulating emotions: An active inference model of emotional state inference and emotion concept learning, 2019)</i>	14

8 BIBLIOGRAFÍA

- Alfonso, B., Vivancos, E., & Botti, V. (2017). Hacia el modelado formal de agentes afectivos en una arquitectura BDI. En *ACM Trans. sobre tecnología de Internet* (Vol. 17, pág. 5).
- Argente, E., Botti, V., Perez-Garcia, D., & Del Val, E. (2020). Normative Emotional Agents: a viewpoint paper. En *IEEE Transactions on Affective Computing* (págs. 1-1).
- Bakhtiyari, K., & Husain, H. (2014). Fuzzy model of dominance emotions in affective computing. En *Neural Computing and Applications* (Vol. 25, págs. 1467–1477).
- Barrett, L. (2017). *How Emotions are Made: The Secret Life of the Brain*. New York: Houghton Mifflin Harcourt.
- Baslet, G., Termini, L., & Herbener, E. (2009). Deficits in emotional awareness in schizophrenia and their relationship with other measures of functioning. *J. Nerv. Mental Dis*(197), 655–660.
- Berthoz, S., Ouhayoun, B., & Parage, N. (s.f.). Etude preliminaire des niveaux de conscience emotionnelle chez des patients deprimes et des controles. *Ann. Med. Psychol*(158), 665–672.
- Boiger, M., Ceulemans, E., De Leersnyder, J., Uchida, Y., Norasakkunkit, V., & Mesquita, B. (2018). Beyond essentialism: Cultural differences in emotions revisited. En *Emotion* (Vol. 18, pág. 1142).
- Botti, V., Carrascosa, C., & Julián, V. (s.f.). Tema 2: Agentes Inteligentes: Conceptos Fundamentales. *Material de Teoría de Agentes Inteligentes*. Recuperado el 25 de 04 de 2022
- Brown, T. H., Zhao, Y., & Leung, V. (2009). Hebbian plasticity A2–squire. En *Encyclopedia of Neuroscience* (págs. 1049–1056). R. Larry.
- Ciarrochi, J., Caputi, P., & Mayer, J. (2003). The distinctiveness and utility of a measure of trait emotional awareness. *Pers. Individ. Differ*(34), 1477–1490.
- Conant, C., & Ashbey, W. (1970). Every good regulator of a system must be a model of that system. En *Int. J. Syst. Sci.* (Vol. 1, págs. 89–97).
- Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. En *Emotion* (Vol. 18, pág. 75).
- Damasio, A. R. (1994). *Descartes' Error : Emotion, Reason, and the Human Brain*. New York.
- Ekman, P. (1992). An argument for basic emotions. En *Cognition & Emotion* (Vol. 6, págs. 169–200).
- Gebhard, P. (2005). ALMA: un modelo de afecto en capas. En *ACM, Proc.gs of the 4th Int. Conf. conjunta sobre Agentes Autónomos y Sistemas Multiagente* (págs. 29-36).
- Hayes, S., & Smith, S. (2005). *Get Out of Your Mind and Into Your Life: The New Acceptance and Commitment Therapy*. New Harbinger.

- Hoemann, K., Xu, F., & Barrett, L. F. (2019). Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. En *Developmental psychology* (Vol. 55, págs. 1830).
- Hohwy, J. (2016). The self-evidencing brain. En *Noûs* (Vol. 50, págs. 259–285).
- Ivonin, L., Chang, H.-M., Chen, W., & Rauterberg, M. (2012). A new representation of emotion in affective computing. En *Proc. of Int. Conf. on affective computing and intelligent interaction* (págs. 337–343).
- Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., . . . Lindquist, K. (2019). Emotion semantics show both cultural variation and universal structure. En *Science* (Vol. 366, págs. 1517–1522).
- Jain, S., & Asawa, K. (2015). EMIA: emotion model for intelligent agent. En *Journal of Intelligent Systems* (Vol. 4, págs. 449–465).
- Jiang, H., Vidal, J. M., & Huhns, M. N. (2007). EBDI: an architecture for emotional agents. En *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems* (págs. 1-3).
- Kazemzadeh, A. (2010). Using interval type-2 fuzzy logic to translate emotion words from Spanish to English. En *IEEE, Int. Conf. on Fuzzy Systems* (págs. 1–8).
- Lackner, J. (2005). Is IBS a problem of emotion dysregulation? Testing the levels of emotional awareness model. *Proceeding of the Presented at the Annual Meeting of the American Psychosomatic Society*.
- Levine, D., Marziali, E., & Hood, J. (1997). Emotion processing in borderline personality disorders. *J. Nerv. Mental Dis*(185), 240–246.
- Lim, N. (2016). Cultural differences in emotion: differences in emotional arousal level between the east and the west. En *Integrative Medicine Research* (Vol. 5, págs. 105 – 109).
- Marsella, S., & Gratch, J. (2009). EMA: un modelo de proceso de dinámica de evaluación. En *Cognitive Systems Research* (Vol. 10, págs. 70-90).
- Mehrabian, A. (1996). “Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. In *Current Psychology* (Vol. 14, pp. 261-292).
- Meuleman, B., & Scherer, K. (2013). Modelado de evaluación no lineal: una aplicación del aprendizaje automático al estudio de la producción de emociones. En *IEEE Trans. sobre computación afectiva* (Vol. 4, págs. 398-411).
- Ministerio de Sanidad. (25 de 04 de 2022). *Bemoción*. Obtenido de <https://bemocion.sanidad.gob.es/emocionEstres/emociones/aspectosEsenciales/qu eSon/home.htm>
- Moeller, S., Konova, A., Parvaz, M., Tomasi, D., Lane, R., & Fort, C. (2014). Functional, structural, and emotional correlates of impaired insight in cocaine addiction. *JAMA Psychiatry*(71), 61–70.
- Panksepp, J., & Biven, L. (2012). *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions*. New York: Norton & Company.

- Real Academia Española. (2022). *Diccionario de la Lengua Española*, 23.^a ed. Obtenido de <https://dle.rae.es/norma?m=form>
- Reisenzein, R. (1994). Pleasure-arousal theory and the intensity. En *Journal of Personality and Social Psychology* (Vol. 67, pág. 525).
- Reisenzein, R., Hudlicka, E., Dastani, M., Gratch, J., Hindriks, K., Lorini, E., & Meyer, J.-J. (2013). Modelado computacional de la emoción: hacia la mejora del intercambio inter e intradisciplinario. En *Trans. IEEE sobre computación afectiva* (Vol. 4, págs. 246-266).
- Russel, J. A., Lwicks, M., & Niit, T. (1989). A cross-cultural study of a circumplex model of affect. En *Journal of Personality and Social Psychology* (Vol. 57, págs. 848-856).
- Russell, J. (1980). Un modelo Circunflejo de afecto. En *Diario de Personalidad y Psicología Social* (Vol. 39, págs. 1161-1178).
- Smith, R., Killgore, W., & Lane, R. (2018). The structure of emotional experience and its relation to trait emotional awareness: a theoretical review. En *Emotion* (págs. 670–692).
- Smith, R., Parr, T., & Friston, K. J. (2019). Simulating emotions: An active inference model of emotional state inference and emotion concept learning. En *Frontiers in Psychology* (pág. 2844).
- Taverner, J., Vivancios, E., & Botti, V. (2020). A multidimensional culturally adapted representation of emotions for affective computational simulation and recognition. En *IEEE Transactions on Affective Computing* (Vol. 546, págs. 1-1).
- Wooldrige, M. (2009). *An Introduction to MultiAgent Systems* (Second ed.). John Wiley & Sons.