



Universidad de Chile

Facultad de Filosofía y Humanidades

Departamento de Filosofía

Adjudicación de estados mentales mediante las emociones

Tesis para optar al Título de Licenciado en Filosofía

Pablo Ignacio Sánchez Zavala

Profesor Guía:

Dr. Manuel Rodríguez Tudor

Santiago de Chile

2019

¿irías a ser muda que dios te dio esos ojos?

(Vicente Huidobro, Altazor, Canto II)

Agradecimientos

Si tuviese que nombrar en un texto a todas las personas y motivos por los cuales estar agradecido, sin los que no hubiese sido posible escribir esta tesis, la verdad sea dicha, aquel texto superaría con creces la extensión de la misma, no obstante, hay menciones especiales que creo no pueden quedar fuera por la misma temática de la tesis, así que tal como en una novela que trate sobre un pescador hay un capítulo que realmente no es de vital importancia para el lector, donde se hace una detallada explicación sobre cómo es que se realiza la pesca y todos los instrumentos usados para la misma y su funcionamiento específicas, considere el lector como una página totalmente saltable, pero parte del texto.

La primera persona y con la que me hayo más en deuda es mi madre Pamela, quien no sólo me ha acompañado durante toda mi vida, sino que desde pequeño me brindó las herramientas para saciar mi sed de conocimiento que desembocó en mi interés por la filosofía. Seguido de ella, menciono a mi abuelo Carlos, quien ha sido un padre para mí y a mi abuela Hilda, cuya paciencia para mis hábitos de trabajo poco convencionales me brindó un lugar tranquilo para trabajar a gusto. También me es preciso mencionar a los padres de mi pareja Jacqueline y Jorge, ya que ellos siempre me brindaron un lugar junto a ellos cuando lo necesité a lo largo de la carrera, pero más especialmente a Constanza, ya que ella siendo mi pareja, ha sido quien más ha contribuido durante la experiencia de crear este texto, y a quien va dedicado de manera muy especial. Me gustaría nombrar a Manuel Rodríguez, quien siendo mi profesor asesor ha tenido una paciencia de oro conmigo y su buen trato es siempre motivante para dar cada paso, además de su lucidez a la hora de despejar mis dudas y sacarme de la selva oscura que puede llegar a ser trabajar temáticas como esta. Además, quiero agradecer al profesor Rodrigo González, quien junto con el profesor Manuel me ha dado muy buenos consejos, tanto para la realización de este trabajo como para mi desarrollo como (perdóneme el lector por mi atrevimiento) filósofo y pensador, teniendo la virtud desde mi mirada de haber sido quien me inspiró por ahí en mi segundo año a decantarme por estos asuntos de la filosofía, y no otros por maravillosos que sean. Finalmente quiero agradecer a todos los que hayan sido y serán mis alumnos (pero especialmente a los que han sido), ya que al ser un sujeto tan afortunado de empezar muy joven mi quehacer pedagógico, he tenido la suerte de guiar a chicos que ninguna palabra menor a “maravillosos” podría describir, y en mis momentos de mayor debilidad han sido ellos quienes sin saberlo me han dado la fuerza para seguir adelante.

Índice

Agradecimientos	3
Resumen	5
Introducción	6
Capítulo 1: Dos problemas siameses	7
Capítulo 2: Dos antecedentes históricos importantes: Descartes y Turing	11
Descartes y Turing.....	11
Descartes	16
Turing y el internalismo	21
Capítulo 3: El debate actual, la ciencia cognitiva y teoría de la mente ToM ..	29
La ciencia cognitiva.....	29
Arquitectura de la mente	31
Teoría de la mente (ToM).....	32
Capítulo 4: Sobre las emociones	38
La distinción entre las emociones y los sentimientos.....	43
Los sentimientos y la conexión entre mente y cuerpo	53
La empatía	56
Phineas Gage, las teorías clásicas y la intersubjetividad.....	61
Capítulo 5: una aproximación emocional al método y el criterio de adjudicación de estados mentales	71
El criterio emocional	71
El método emocional.....	73
Conclusiones.....	77
Bibliografía.....	78

Resumen

La presente tesis tiene como propósito analizar el rol que cumplen las emociones al momento de reconocer en otros individuos estados mentales, para esto, durante todo el texto se propone si es posible probar que la posesión de emociones implica que exista una mente en la que puedan contenerse. En un principio se enfatiza en el criterio y método para adjudicar estados mentales, luego en el pensamiento de René Descartes sobre la mente y cómo es posible identificar si el otro la tiene. Para contrastar la idea del filósofo francés se considera a Turing y el test que lleva su mismo apellido, cuya intención es cuestionar la adjudicación de estados mentales en máquinas. También se toma en consideración la ciencia cognitiva, puesto que sus estudios fueron los primeros en poner este tema filosófico en la palestra científica durante el siglo XX. Posteriormente se explica la diferencia entre emociones y sentimientos, considerando también a la empatía como un agente importante dentro del tema abordado. Para relacionar de mejor manera al lector con estos tres últimos puntos y para fines de desarrollo del mismo trabajo, se utilizan como ejemplos por un lado el videoclip “I hope your happy” de la banda Estadounidense Blue October, donde es posible apreciar un autómata (interpretado por una actriz) el cuál permite desarrollar el problema acerca de la adjudicación de estados mentales en seres que no están vivos, y por otro lado el caso de Phineas Gage, un obrero que a causa de un accidente, su vida en el aspecto social cambió rotundamente, que en contraste con el ejemplo anterior, debido a que es un caso real el cual le sucedió a una persona, permite tener un amplio espectro del problema principal del trabajo. Teniendo todos estos antecedentes, el texto finaliza con el desarrollo, análisis y conclusiones del propio autor de este.

Palabras clave: Estados mentales; Emociones; Emocionalidad; Mente.

Introducción

Piense el lector en alguna ocasión que le hayan dicho “se más racional”, “cabeza fría, corazón caliente”. Este tipo de frases esconden una concepción sobre la racionalidad como separada de la emocionalidad. La razón se concibe como un proceso lógico, llevado por un entendimiento de los procesos de causa y efecto, mientras la emocionalidad es aquello que se siente y nubla el proceso racional debido a la volición de quien la experimenta.

Aún ante lo anteriormente dicho, es difícil detectar un momento en la vida de un individuo en que este no presente emociones. Por tanto, es interesante revisar el rol que cumplen las emociones en distintos procesos cognitivos.

En este trabajo pretendo proponer una alternativa ante algunas propuestas clásicas sobre un problema filosófico específico “*el problema de las otras mentes*” que consta de hacer la pregunta ¿Qué nos faculta a pensar que los otros tienen mentes? La tradición dicta que existen dos alternativas para responder esta pregunta, la primera es la posesión de lenguaje, que consta de una capacidad que tiene el individuo para entender el lenguaje tanto a un nivel gramatical y semántico, por otro lado, existe el criterio de “comportamiento inteligente”. Sin intentar validar estos criterios, propongo uno más a la ecuación: la “emocionalidad”, en tanto un ser que sea capaz de experimentar emociones debe tener estados mentales. Además, propongo que el mejor método para saber si es que otro tiene estados mentales son también las propias emociones, posicionándome desde una postura de primera persona “internalista”.

Este trabajo tiene un carácter puramente especulativo, y si he mencionado trabajos ajenos con un carácter científico, es solo como datos a considerar. Aclaro que no pretendo presentar ninguna arquitectura mental propia ni adherir a alguna de las posturas de ciencia cognitiva que expondré a lo largo del texto, sin embargo, me serviré principalmente de los trabajos de Antonio Damasio, neurobiólogo dedicado al estudio de las emociones y sentimientos a un nivel neurológico, el cuál describe la emocionalidad de modo bastante claro.

Por último, trataré como experimento mental la posibilidad de la inteligencia artificial, pues no pretendo que ninguna de las propuestas aquí presentes niegue la posibilidad de existencia de estados mentales en seres no humanos.

Capítulo 1: Dos problemas siameses

La presente tesis tiene como propósito realizar un rescate a la emocionalidad, ya que esta ha sido dejada históricamente de lado por los filósofos, y en caso de tratarla, rara vez ha sido el propósito fundamental de su obra, a pesar de que esta es parte constituyente de la condición humana. Ver los límites de la influencia de la emocionalidad sería mucho más de lo que se me permitiría para un trabajo, después de todo, a una edad adulta ya es difícil encontrar algo que no estimule algún tipo de emoción y una empresa de ese tipo sería tan extensa que no cabría en una tesis y quizá tampoco en una vida. Por lo cual en un principio decidí enfocarme en la teoría del conocimiento y un posible lazo con las emociones; pero esta seguía siendo una empresa imposible de llevar a cabo para mí, por lo cual decidí centrarme en un aspecto en específico, a saber, las relaciones humanas, pero sin renunciar a mis pretensiones iniciales de trabajar teoría del conocimiento, decidí intentar trabajar la pregunta ¿Podemos llegar realmente a conocer a alguien? Sin embargo, esta sigue estando fuera de los límites prácticos a los que me enfrento, ya que para contestar eso primero se debe considerar ¿Qué es el otro?, ¿Puedo adjudicarles estados mentales?, ¿Qué es conocimiento? entre otras preguntas y problemas que han rodeado la historia de la filosofía. En este contexto, ya habiendo decidido trabajar filosofía de la mente, pues ahí es donde las emociones podrían presentarse de mejor manera, ya que parece que la mente y la emocionalidad tienen una relación estrecha, me centré en un problema en específico, a saber, el *problema de las otras mentes*. Por tanto, me propuse investigar si es que la emocionalidad tendría algo que ver con cómo es que se adjudican estados mentales, es ahí donde reflexionando en mí fuero interno contrasté el origen del problema, el cual se presenta con la figura de René Descartes. Una buena explicación de esto es la expuesta por G. Ryle en la siguiente cita:

Se da así una oposición entre mente y materia que, a menudo, se describe de la siguiente manera. Los objetos materiales se encuentran ubicados en un campo común, el "espacio", y lo que acaece a un cuerpo está conectado mecánicamente con lo que les sucede a otros cuerpos ubicados en otras partes del espacio. Pero los hechos mentales acaecen en ámbitos aislados, las "mentes", y no existe conexión causal directa entre lo que le sucede a una mente y lo que le pasa a otra, con excepción, quizá, de la telepatía. La mente de una persona puede afectar la mente de otra únicamente a través

del mundo físico. La mente es su propio espacio y cada uno de nosotros vive la vida de un fantasmal Robinson Crusoe. Podemos vernos, oírnos y empujarnos los unos a los otros, pero somos irremediablemente ciegos, sordos e inoperantes con respecto a la mente de los demás. ¿Qué tipo de conocimiento puede obtenerse de las operaciones de una mente? De acuerdo con la doctrina oficial, toda persona tiene un conocimiento inmejorable de dichas operaciones. Los estados y procesos mentales son (o lo son normalmente) estados y procesos conscientes que no engendran ilusiones ni dan lugar a dudas. Los pensamientos, sentimientos y deseos de una persona, sus percepciones, recuerdos e imágenes son intrínsecamente "fosforescentes"; su existencia y naturaleza son reveladas inevitablemente a su dueño. La vida interior es una corriente de conciencia de tal tipo que sería absurdo pretender que la mente —cuya vida es una corriente— pudiera ignorar lo que le está pasando. (Ryle, 2005)

Frente a este problema filosófico me encontré con que las emociones podrían ser una buena respuesta, si es que se puede probar que la posesión de emociones implica que exista una mente en la que puedan contenerse, la posesión de emociones sería un criterio suficiente para asumir que el otro tiene una mente. Sin embargo, esto presenta otra dificultad, aún si pruebo que el otro tiene mente porque tiene emociones, entonces ¿Cómo sé que el otro tiene emociones? La respuesta a esta pregunta es la misma que la primera, “la emocionalidad funciona como un buen método”, si se considera el concepto de “empatía” que suele dar a entender que un sujeto “siente” o experimenta lo mismo que siente el otro, abre una alternativa, ya que la mente de algún modo sí encuentra en sí misma la información que debiese quedar privada para los demás, he aquí el motivo por el cual he decidido llamar a este primer capítulo “problemas siameses” ya que el problema de las otras mentes se puede dividir en dos problemas distintos: 1) ¿Qué criterio es usado para adjudicar estados mentales?; y 2) ¿Qué método es usado para aplicar aquel criterio? Unidas por un fenómeno en común (emociones).

Creo que la respuesta a la primera pregunta sería “si un ser en particular experimenta emociones, entonces, debe tener una mente” y para la segunda “en la medida en que se presentan reacciones empáticas, experimentando las emociones del otro, se tiene un método para asumir la presencia de estados mentales”. Esto se sitúa en una vereda diferente a la tradición que considera como criterio y metodología el comportamiento inteligente y la

posesión de lenguaje, no obstante, no busco invalidar las posturas clásicas en tanto a la utilidad que estas presentan, por el contrario, sigo considerando que abordar el problema desde las perspectivas clásicas es válido en la medida que no implique dejar de lado la emocionalidad.

No obstante, para iniciar este proceso, primero se debe revisar el pasado en busca de pistas que describan parte del desarrollo de esta discusión, para lo cual tomaré a dos pensadores en particular: René Descartes y Alan Turing. Al primero por ser quien da origen a la discusión, tópico que lo pone en el centro de la esfera, como también por sus contribuciones a la tradición, por otro lado la figura de Turing fue una decisión más compleja, claramente el antes que filósofo fue matemático, y su trabajo tenía un enfoque más científico que filosófico, tomando una gran distancia de Descartes, pero lo elegí a él porque si se lo trata de una manera laxa (sobre especulando un poco), la teoría de la mente que este “propone” resulta beneficiosa a la hora de trabajar mi teoría, pues no pretendo dar a entender un lazo entre el ser humano y las propiedades mentales, considerando tanto a los animales, extraterrestres y máquinas como capaces de tener una mente; sin embargo, no entraré en discusiones sobre si los animales y extraterrestres tienen estados mentales. Ahora bien, dicho esto trabajaré los estados mentales en máquinas, pero tampoco pretendo realizar un trabajo extenso sobre la inteligencia artificial, y la trataré más desde una perspectiva de “ciencia ficción” que científica cognitiva, es por ello que la propuesta de Turing me resulta útil en la medida que se defiende la posibilidad de inteligencia artificial y un método para evaluarla igualmente válido para seres humanos.

Es en este punto que se vuelve necesario explicar qué son las emociones, los sentimientos y la empatía, ya que sin ellos no se puede llevar a cabo una tesis que trate de la adjudicación de estados mentales mediante las emociones. Para aclarar las cualidades y explicar el funcionamiento de estas y su rol en la temática, me serviré de los trabajos y la postura de Antonio Damasio, quien se ha dedicado al funcionamiento del cerebro y cómo es que las emociones se presentan en este órgano. Si bien los trabajos de Damasio me resultan muy útiles, yo no tengo ningún compromiso decisivo con su teoría, por lo cual la adaptaré en los

aspectos relevantes a un sistema propio, siempre dejando en claro al lector el momento en que realizaré estas adaptaciones. En segundo lugar, como ya he dicho, Damasio es un neurocientífico, por lo cual sus exposiciones constan principalmente de explicar el funcionamiento del cerebro, en cambio yo trabajo desde un campo especulativo y no científico, por lo cual tampoco realizaré descripciones detalladas sobre el cerebro o el organismo del ser humano más allá de lo que me resulte necesario.

Aun así, antes de la explicación de la postura de Damasio sobre la emocionalidad, creo necesario dedicar un capítulo a la ciencia cognitiva, (que se establece como un tipo de estudio interdisciplinario dedicado a estudiar la mente), ya que inevitablemente este trabajo está directamente relacionado con esta corriente, por lo que sería imposible tratar el tema sin ponerse en contexto de las sub-corrientes que presenta, puesto que son importantes para dar mi propia perspectiva sobre el tema posteriormente. De antemano aviso al lector que no he asumido ningún compromiso previo con ninguna de estas sub-corrientes, y su exposición en el capítulo tres será meramente explicativa, para ya en el capítulo cuatro realizar una crítica de estas y finalmente explicar mi postura personal en el capítulo cinco, tras un análisis de los casos de Phineas Gage y un experimento mental, sacado del videoclip de la banda Blue October para la canción I hope you're happy; si bien expongo los motivos teóricos por los cuales este material en específico es útil durante el capítulo cuatro, existe un motivo por el cual trabajar una fuente tan poco académica, sosteniéndome en que, al tratarse de emocionalidad, sería mejor trabajar una obra artística que despierte las sensaciones del lector, y ante la gran capacidad de análisis que ofrece la filosofía desde la cual se puede tomar una perspectiva frente a casi cualquier temática, un análisis de este video no me pareció descabellado, en cambio resulta una apuesta didáctica y pedagógica al poner un ejemplo tan cotidiano como un video.

Capítulo 2: Dos antecedentes históricos importantes: Descartes y Turing

Descartes y Turing

Si bien el estudio sobre la mente se remonta a la antigua Grecia, con la diferenciación entre “nous” (intelecto) y “psike” (alma), pasando por toda la filosofía medieval, realmente no se conformó de manera seria e independiente hasta llegar a la modernidad con el personaje René Descartes, como describe John Searle en su libro “La mente, una breve introducción” (2006):

En la era moderna, la filosofía de la mente comienza en efecto con la obra de René Descartes. Descartes no fue el primero en sostener puntos de vista como los que sostuvo, pero su concepción de la mente fue la más influyente entre las propuestas de los llamados filósofos modernos, los filósofos del siglo XVII y siguientes (Searle, 2006)

El lenguaje que se utiliza para el estudio en la filosofía de la mente es imperantemente Cartesiano y fundado en el *punto arquimédico* de su metafísica. La fórmula *cogito ergo sum*, que se puede traducir por “pienso por lo tanto existo”¹, deriva en lo que René Descartes llama *res cogitans* o en español “cosa pensante”; esta concepción está cimentada en la cosmovisión cristiana sobre el “alma” a la que Descartes adhería. A pesar de que el filósofo francés se tomó su tiempo para poder probar este punto, (argumentación que será expuesta más adelante en el texto) lo cierto es que siempre tuvo como presupuesto que el “alma” debe ser inmortal y diferente al cuerpo; a esta postura filosófica se le llama “dualismo de substancia” o “dualismo substancial”.

En el dualismo substancial se presentan varios problemas, los cuales han quedado como legado de Descartes para los filósofos posteriores, entre ellos está el problema abordado dentro de este texto llamado comúnmente “el problema de las otras mentes”. La pregunta, al menos de manera muy sucinta es ¿Cómo sabemos que los otros realmente tienen mentes?,

¹ He traducido *ergo* por “por lo tanto”. Sin embargo, la traducción literal es “luego”. Lo he traducido de esta manera, ya que en el español la palabra “luego” se presta para una diferenciación en tanto al tiempo, y el hecho de si “existimos” después de pensar o mientras pensamos es un debate irrelevante para este texto.

teniendo en cuenta que en un lenguaje más cartesiano, es posible reemplazar el concepto de mente por el de “pensamiento” o “alma”².

Descartes plantea la pregunta con el ejemplo de los “autómatas” o los “hombres de las capas y sombreros”, presente tanto en las *Meditaciones metafísicas* como en *el Discurso del método*. El experimento mental consta de preguntarse si los “seres” que transitan al otro lado de la ventana realmente son hombres dotados de pensamiento, o si en realidad son máquinas o “robots” que fueron vestidos de tal manera que no se pudiera observar sus rasgos y dar cuenta de su real composición material. El planteamiento va más allá de lo meramente textual y expone el hecho de que si se acepta el paradigma dualista, habría que enfrentarse a una incapacidad para determinar si es que el otro realmente tiene o no estados mentales, dicho de otra manera, al encontrarse una persona X con una persona Y, X solo verá un cuerpo, además X podría oler a Y o escuchar lo que tenga que decir, y si está hablando en su mismo idioma, quizá entenderlo; también podría tocar a Y, y si se quiere incluso saborearlo. Sin embargo, lo único con lo que se topará serán sus propios estados mentales, él comprenderá las palabras y por su mente pasará el perfume que, haya elegido ponerse aquella mañana, pero nada de eso se asemejará a los pensamientos que tenga Y en aquel instante, puesto que según Descartes sólo tenemos sensaciones de las cosas materiales, y como el cuerpo es diferente al pensamiento, el pensamiento del otro se vuelve inalcanzable mediante las capacidades perceptuales. En esto juega las capacidades introspectivas de la substancia mental, el pensar en sí misma es parte constitutiva de ella, ya que es esencialmente pensamiento, en cambio las cosas materiales deben ser pensadas fuera de sí mismas y por ello están siempre en una esfera indirecta a la substancia mental, Ryle lo explica así:

Además de otorgarse a toda persona la posesión de estos datos inmediatos de la conciencia, también se supone que es capaz de ejercitar de tiempo en tiempo un tipo especial de percepción: la interna o introspección. Toda persona puede echar una "mirada" (no óptica) a lo que está pasando en su mente. No sólo puede mirar y escudriñar una flor mediante el sentido de la vista y escuchar y discriminar las notas de una campana mediante el sentido del oído; también puede, reflexiva o introspectivamente, observar los episodios corrientes de su vida interna sin el concurso de órganos corporales de los sentidos. También se supone comúnmente que esta autoobservación es inmune a la ilusión, confusión o duda. Los informes de

² Introduzco este comentario puesto que me tomaré la libertad de usar estos términos como sinónimos e intercambiarlos entre sí a lo largo de este.

una mente acerca de sus propios estados tienen una certidumbre superior a la mayor que puedan poseer los informes referentes a cuestiones del mundo físico. Las percepciones sensoriales pueden ser erróneas o confusas, pero la autoconciencia e introspección no. (Ryle, 2005)

El caso del lenguaje es particular, ya que Descartes lo considera como una prueba de pensamiento. A pesar de que en su época no existía algo así como un “chatbot”³, él propone que aunque una máquina fuese hecha para emitir palabras en un idioma particular, esta nunca podría llegar a obtener la elocuencia de una persona, en otras palabras, esta carecería de aquello que le permitiría considerarse como “hablante competente” (la capacidad de entender ironías, metáforas, eufemismos, etcétera), además, Descartes lleva el ejemplo más allá sacando a colación el caso de ciertos tipos de loros, los cuáles a pesar de ser capaces de emitir palabras, de todas formas carecen de un lenguaje competente.

Si bien René Descartes también propone como criterio el comportamiento inteligente⁴, el criterio lingüístico parece ser mayor, puesto que el comportamiento se manifiesta en el cuerpo, el cual no está dotado por sí mismo de mente y podría tratarse de un autómeta.

En la misma línea de Descartes, pero ya varios siglos después, encontramos los trabajos de Alan Turing. Si bien hubo muchos filósofos que trataron el problema en el lapso entre los dos pensadores, lo cierto es que Alan Turing es fundamental a la hora de tratar el problema de las otras mentes⁵; antes de él realmente ningún filósofo excepto los solipsistas se atrevían a negar la presencia de estados mentales en hombres, y a pesar de que Turing no los negó, abrió la puerta para pensar en seres con instanciaciones materiales no humanos dotados de

³ Los chatbot son programas con los que, se supone, uno puede “conversar”.

⁴ El comportamiento inteligente según lo explica Ryle en *el concepto de lo mental* sería una especie de comportamiento similar al propio de tal manera que se pueda realizar una analogía entre uno mismo y el otro, para inducir que el otro al igual que yo tiene estados mentales. Aunque posteriormente en crítica al dualismo de substancia, que implica un error categoría (el pensar a ambas substancias como similares) agrega que sería necesaria la comparación entre varios individuos para generar un concepto de comportamiento inteligente.

⁵ Turing en realidad no se centró en el problema de otras mentes, sino que exploró un posible método para evaluar si es que las máquinas pueden ser inteligentes. No obstante, por su contra postura a Descartes (implícita), trató a Turing y su propuesta del juego de la imitación como un tratamiento al problema de las otras mentes

mente⁶. Alan Turing es considerado como el padre de la filosofía de la inteligencia artificial, la cual se dedica puntualmente a la pregunta ¿pueden las máquinas pensar?

A la época del matemático, ya se había abandonado el paradigma dualista de Descartes, y los pocos que seguían defendiéndolo (al menos según Searle) lo hacían más por compromisos religiosos que por un examen filosófico real, además, si a esto se le suman los grandes avances que se dieron durante el periodo entre Descartes y Turing en el campo de la medicina, el cerebro tomó un rol fundamental, dando paso a inclinaciones fuertemente materialistas⁷; en esta atmósfera es que Turing propone el problema de la inteligencia artificial, problematizando sobre el punto de si es lo mismo un cerebro vivo que una mente.

Turing fue un fiel creyente de la posibilidad de crear una inteligencia artificial IA, el Matemático vio en los computadores digitales un futuro para el proyecto, pero se le presentaba como problemática lo arcaico de estos para su época, así que decidió fijar un criterio para la presencia de estados mentales en máquinas, el cuál fue llamado “el juego de la imitación” o “test de Turing” como se lo conoce popularmente.

Rodrigo González explica el Test en su texto: *El Test de Turing: dos mitos, un dogma* de la siguiente forma:

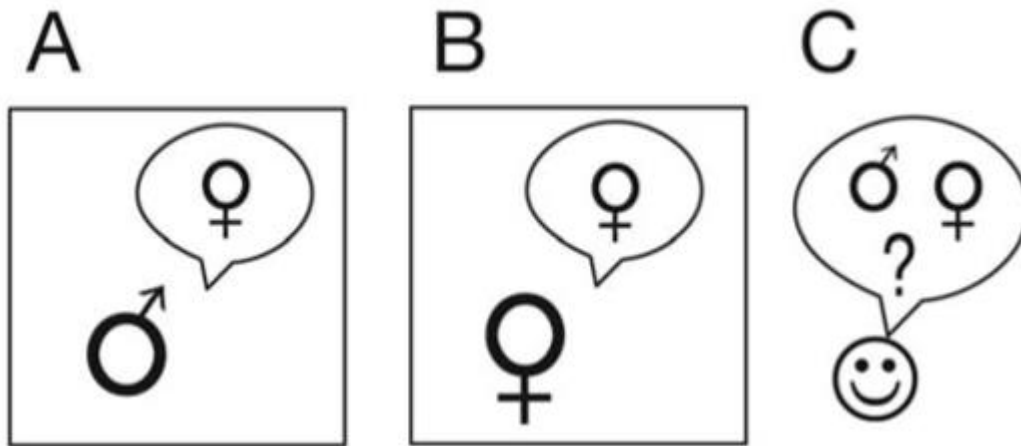
Turing planteó diversas versiones del Juego de la Imitación. En la original, tres personas participan en el juego: un hombre, una mujer y una persona cuyo sexo no es relevante. Tanto A (hombre) como B (mujer), quienes se encuentran en dos piezas distintas, responden a las preguntas formuladas por C (‘juez’ de sexo no relevante), quien debe identificar el sexo de los otros participantes desde el exterior. Mientras que la mujer responde a las preguntas de manera veraz, el hombre debe hacerse pasar por una mujer, de modo que C crea que A también es de sexo femenino.

Para evitar que se descubra de manera obvia quién es quién, las preguntas y respuestas son escritas e impresas a través de tele-impresoras. La figura 1 esquematiza la versión original del Juego de la Imitación:

⁶ Si bien también se puede considerar el problema sobre si los animales tienen mente dentro de la discusión, he decidido dejarlo fuera por motivos meramente prácticos para el presente trabajo.

⁷ Los materialistas en resumidas cuentas creían que la mente y el cerebro eran lo mismo, y por ello cualquier cosa que no tuviese cerebro no podía pensar.

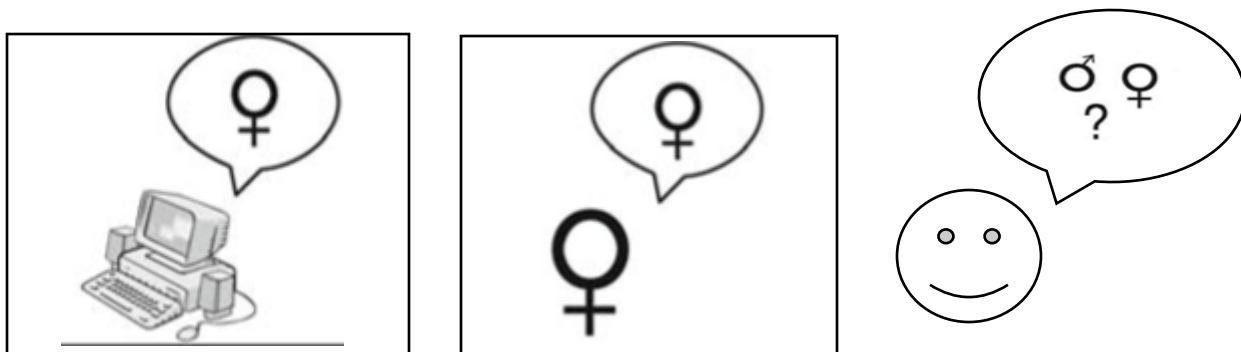
Fig. 1



Si A es lo suficientemente convincente, C concluirá que hay dos mujeres en vez de una, lo cual revela el peculiar papel que A desempeña, a saber, debe causar que C se equivoque en la identificación. A juicio de Turing, lo anterior indica cómo el juego puede reemplazar la pregunta acerca de si las máquinas piensan. Si un computador reemplaza a A, y ello no involucra ninguna diferencia respecto de cómo C identifica erróneamente a los participantes, entonces el computador es capaz de generar comportamiento inteligente, puesto que puede suplantar eficazmente a un humano.

La segunda versión del juego consiste justamente en que un computador reemplaza a A, haciendo creer a C que es una mujer. B, en tanto, sigue respondiendo verazmente, tal como ilustra la figura 2:

Fig. 2



Si el computador responde en forma eficiente a las preguntas que se le formulan, el resultado más probable es que C se equivoque y concluya que hay dos mujeres en vez de una. (González, *El Test de Turing: Dos mitos, un dogma*, 2007)

El *juego de la imitación* demuestra el mayor punto de conexión entre el autor inglés y el filósofo francés, y es que se considera al ser pensante como dotado de un lenguaje, pero a su vez cabe notar que en ninguno de los dos casos el intelecto se ve reducido al lenguaje, sino, más bien se ve que la posesión de lenguaje es un indicador de que existe un intelecto. Finalmente, ante lo somero que resultan estas exposiciones, es necesario realizar una exposición un tanto más detallada de las posturas de ambos filósofos.

Descartes

1) El *cogito*:

Resulta imposible referirse a la filosofía de la mente de René Descartes sin tratar primero el concepto de *cogito* contenido en *las meditaciones metafísicas*. En la época que Descartes escribió aquel libro se dio a su vez una reaparición del escepticismo pirrónico, el cual propone que realmente el ser humano no puede referirse a algún hecho del mundo sin tener un riesgo de equivocarse, por lo tanto, todo resulta dudable. Descartes negándose a esta concepción gnoseológica, decide emprender la industria de rearmar el árbol de la metafísica, que según él era el sustento para las ciencias. Él creía que, si se tienen presupuestos metafísicos firmes, las ciencias naturales no podrían fallar. El autor francés consideró que aquella firmeza se encontraba en un modelo como el de las matemáticas, según el cual todas las aseveraciones están respaldadas y siguen caminos tales que cualquier error es meramente metodológico y no en su sustento mismo. Si se falla en sumas como que dos más dos es cuatro, o que la suma de los ángulos internos de un triángulo es igual a dos ángulos rectos, no se habrá encontrado una raíz en la certeza de las conjeturas, sino en el orden. Del mismo modo es necesario encontrar un punto de absoluta certeza en la metafísica, por lo cual Descartes parte las meditaciones desechando todo lo que había tenido por cierto sin alguna prueba fehaciente, lo cual le implica descartar toda la información brindada por los sentidos, su cuerpo mismo e incluso sus conocimientos matemáticos.

Ya en la segunda meditación Descartes se da cuenta de que no puede dudar de él mismo: “¿no me he persuadido de alguna manera de que no existo? No, de ninguna manera; si me he persuadido o, simplemente, si he pensado algo es que era.” (Descartes, *Meditaciones metafísicas*, 2011), pero aquello le presenta la siguiente duda, ¿si es que él *es*, entonces que cosa es?, para contestar, el filósofo moderno hace un recopilatorio de todas las cosas que ya

ha puesto en duda y da por descartado el cuerpo y los sentidos dejando como suyo el pensamiento y nada más: “(...) pensar, y aquí sí encuentro que el pensamiento es un atributo que me pertenece: es el único que no puede separarse de mí. Yo soy, yo existo, esto es cierto; pero ¿cuánto tiempo? Pues todo el que dure mi pensar” (Descartes, *Meditaciones metafísicas*, 2011) .

Aún después de esto, Descartes no se queda tranquilo e inicia una enumeración de los atributos que implica ser una cosa que piensa. Atributos que resultan no menores para la filosofía de la mente, pues a estos posteriormente se les llamará “estados mentales”, por tanto, un ser que posea las siguientes cualidades es un ser “pensante”: “así bien, ¿Qué soy entonces? Una cosa que piensa. ¿Qué es una cosa que piensa? Pues una cosa que duda, que conoce, que afirma, que niega, que quiere, que no quiere, que imagina también y que siente” (Descartes). Para toda la filosofía de la mente posterior, estas cualidades se consideran como necesarias, no así suficiente por sí mismas, sino, en conjunto. Pese a eso yo he decidido sumar a estas (aún contra Descartes) que debe ser capaz de “emocionarse”, por motivos que serán abordados durante el capítulo cuatro.

2) *Dualismo de substancias:*

Ya visto qué significa el pensamiento para Descartes, cabe preguntarse ¿qué es el cogito? En la fórmula pienso luego existo, se da a entender una relación entre el pensamiento y la existencia de tal manera que uno se sigue del otro, en otras palabras, la existencia según Descartes es sinónima del pensar y el pensar a su vez es prueba fehaciente de existencia. Descartes considera que la cosa que piensa o res cogitans es una substancia, pero ¿qué es una substancia? Descartes considera que las substancias son aquellas cosas “claras y distintas” o de manera más simple, objetos que existen con relativa independencia de otros (esta independencia no incluye a Dios). Aparte de la res cogitans, el francés considera otro tipo de substancias las cuales él llamo res extensa, éstas constituyen a todos los objetos que poseen múltiples divisiones, y por tanto son susceptibles de destruirse, a diferencia de la res cogitans que se compone de modo inmaterial. El ser humano por tanto sería un compuesto de *res cogitans* (pensamiento) y *res extensa* (cuerpo), como lo grafica la siguiente tabla:

Fig. 3

	Sustancias	
	Mente	Cuerpo
Esencia	Pensamiento (conciencia)	Extensión (posee dimensiones espaciales)
Propiedades	Conocida directamente Libre Indivisible Indestructible	Conocido indirectamente Determinado Infinitamente divisible Destructible

(Searle, 2006)

La tabla grafica las diferencias entre una sustancia y otra, además agrega el factor gnoseológico que implica el dualismo de sustancia.

Ya que en este sistema son dos sustancias las que tienen que coexistir, es necesario ver el punto en el cual estas dos se unen para conformar al ser humano en particular. Tal como Descartes lo piensa, el pensamiento habita el cuerpo de una manera muy unida, él da el ejemplo de la disanalogía del navío y el capitán: *“no sólo me hospedo en mi cuerpo como un piloto en su navío, sino, que además de eso, estoy unido a él muy estrechamente y de tal forma confundido y mezclado que compongo con él como un solo todo”* (Descartes 2011).

A pesar de que estén unidas ambas sustancias, existe una diferencia entre las maneras de conocer los datos de ambos; el conocimiento de la *res cogitans* es directo, ya que es una inspección de ella misma, en cambio el conocimiento de la *res extensa* se presenta de manera indirecta, pues las dolencias que presente esta deben manifestarse en la *res cogitans* de alguna manera. Descartes da de ejemplo los dolores, si el cuerpo sufre algún daño, este conocimiento se presenta en la mente a manera de dolor, del mismo modo que los datos de los sentidos como las imágenes proporcionadas por la vista, se manifiestan en el pensamiento de manera distinta según como estén posicionados los receptores de los sentidos, aún para un mismo objeto. Finalmente, cabe decir que estas interacciones se justifican en la creencia de Descartes en que el cuerpo y el alma se unen en la glándula pineal, ya que es la única zona del cerebro (que a su vez es el destino de todos los nervios, quienes se encargan del movimiento del cuerpo) que no tiene un símil o es “única”.

3) La adjudicación de estados mentales:

Es necesario repasar el problema de las otras mentes en René Descartes. Como ya se ha dicho antes, el problema de las otras mentes se presenta debido al dualismo de substancias; ya que el único conocimiento directo que se tiene es la propia mente, habría que graduar los conocimientos indirectos, a saber, los cuerpos externos y las otras mentes. Si es que el cuerpo humano es una “máquina” piloteada por el intelecto, el cuerpo debería reflejar al menos de manera parcial los estados mentales. Si se encuentran dos personas y la primera tratase de tener datos sobre la mente de la segunda, para empezar, se toparía con una incapacidad para tener percepciones sobre la otra mente, ya que los únicos datos sensoriales que recibe son de otro cuerpo, y aunque tomase en cuenta que puede utilizar el conocimiento que obtiene de su propia mente, ya que esta le resulta inmediata de conocer (cualquier intento de conocimiento implica pensar y el pensamiento es la propia *res cogitans*, por tanto no hay mediadores en este conocimiento) y de esta forma intentase mediante su propia mente generar algún tipo de prueba se vería truncado, pues lo único que encontrará será su propia mente. Si bien en la segunda meditación (casi al final después del ejemplo de la cera) Descartes vislumbra el ejemplo de los autómatas, no es hasta la quinta parte del *discurso del método* que el francés plantea una solución al problema:

Al llegar a este punto, me detuve muy especialmente en mostrar que, si hubiera máquinas que tuviesen los órganos y la figura exterior de un mono o cualquiera otro animal irracional, no tendríamos ningún medio de reconocer que no era en todo de igual naturaleza que estos animales, al paso que, si hubiera otras semejantes a nuestros y que imitasen nuestras acciones (...), siempre tendríamos dos medios seguros de reconocer que no son hombres. El primero sería que jamás podrían usar de las palabras ni de otros signos compuestos de ellas como lo hacemos nosotros para declarar nuestros pensamientos (...) consiste el segundo medio en que, por más que estas máquinas hicieran muchas cosas tan bien o acaso mejor que nosotros, infaliblemente se equivocarían en otras, y así se descubriría que no actúan por conocimiento (Descartes, Discurso del método, 2001)

Las implicaciones filosóficas de este fragmento, aunque sea breve, son muy profundas; en primer lugar, Descartes maneja como presupuesto “solo los humanos son seres pensantes”, aunque en la argumentación se nota que esto no se debe a la falta de visión que proporcionaba

la limitada tecnología de la época, ya que él propone que una máquina podría adaptarse para lucir exactamente como un ser humano, pero su límite estaría sujeto a su composición material⁸ (carente de alma), y es aquí donde sí se vuelven evidentes las limitaciones que le proporcionó vivir en su tiempo, pero ¿quién podría culparlo por no imaginarse que 300 años después habría computadores digitales? a tal punto que la composición del cuerpo ajeno hoy en día no resulta ser un problema; nada imposibilita realmente crear un maniquí con un sistema de articulación y equilibrio similar al de un ser humano, y si no ha sido creado un robot tal que nadie dude de que este es pensante, se debe más a las limitantes en la programación de este, que en la construcción material del mismo, sin embargo, repasaré más aristas de este problema en el sub capítulo siguiente que versa sobre Turing.

Por otro lado, la posición que asume Descartes sobre el lenguaje es un tanto más interesante que la ya expresada. Cuando Descartes habla sobre el uso de las palabras, se refiere principalmente a la semántica (capacidad de interpretar signos de un lenguaje) propia del ser humano: por ejemplo, en Chile se utiliza de manera regular la palabra “tirar” como un eufemismo para referirse a tener relaciones sexuales, lo más probable es que un ser humano con un poco de práctica logre adecuar esta palabra a su lenguaje cotidiano y dotarla del significado correcto en el momento indicado, pero una máquina difícilmente lograría

entender las diferencias acordes al contexto, aunque logre reconocer que el uso sintáctico de la palabra sea el correcto.

De todas maneras, decir que Descartes cree que el lenguaje es igual al pensamiento, sería ir mucho más allá de lo permitido por el texto, Descartes solo llega a afirmar que los seres pensantes tienen lenguaje porque son pensantes, y no que sean pensantes porque tienen lenguaje.

El método cartesiano está fundamentado principalmente en la inspección de los hechos de una manera reflexiva, donde es el mismo intelecto repasa los datos que tiene y delibera sobre la veracidad en tanto no hay espacio a la duda, así la “ecuación” quedaría de la siguiente manera:

⁸ En la época de Descartes la mayor parte de la tecnología era meramente “mecanicista” lo que significa que su operatividad dependía solo de poleas y palancas, y no era posible emular de manera efectiva organismos más complejos.

*Input sensorial*⁹ \Rightarrow *reflexión interna* \Rightarrow *adjudicación de estado mental*

Pero teniendo lo anterior en claro, esta noción resulta cuando menos problemática; si bien es cierto que el hecho de adjudicar o no estados mentales puede ser considerado como una decisión, lo cierto es que esta decisión rara vez es consciente, o por lo menos rara vez lo es en la naturaleza. Cuando un sujeto x se encuentra con otro en la calle, no se detiene a observarlo, escucharlo y pensar ¿es un ser pensante? en cambio, lo más probable es que cualquier ser humano simplemente asuma que el objeto que se encuentra frente de si es otro ser humano pensante, y aunque la duda le atravesase por la mente, es poco probable que estando frente a otro ser humano pueda convencerse a sí mismo de que aquello que se encuentra frente de si no posee estados mentales propios. Pero ¿qué pasa con las condiciones actuales? Hoy en día son múltiples las instancias donde podríamos hablar con algo que no esté dotado de una mente, por ejemplo, está el caso del asistente de voz que tienen los teléfonos o los GPS que dan indicaciones, intentar encontrar soluciones para estos casos en la filosofía cartesiana sería probablemente insatisfactorio, ya que el asistente de google responde acorde a la pregunta que se le realice de la mejor manera posible e incluso es programable para que lo haga con lucidez aprendiendo los modismos del usuario y el GPS se logra adaptar sin mayores problemas al tráfico que hay en la ruta como lo haría casi cualquier persona con esa información previa, siendo ambos relativamente capaces de cumplir con las condiciones de comportamiento y uso del lenguaje.

Turing y el internalismo

Si bien la importancia de Alan Turing es indiscutible para la historia del siglo XX, si se puede poner en duda la relevancia que tenga para el presente estudio. La importancia radica en el mismo “juego” de la imitación, pero el centro de atención está en el juez que debe evaluar a los participantes. El criterio es el mismo que se presenta en Descartes, el juez debe tomar el input sensorial, y tomar sus decisiones en base a la información. Ya con Descartes había dicho que una situación en la que quepa dudar sobre la actividad mental de otro ser debería

⁹ lenguaje bien articulado y dotado de semántica o comportamiento inteligente.

ser anómala, caso que se cumple perfectamente en el juego, el juez se ve obligado a decidir solo con las respuestas que le den los participantes. El mismo juego de la imitación propone esta situación anómala para dejar en igualdad de condiciones a la máquina y al ser humano, a este criterio se le llama *internalista*, este tiene la ventaja de dejar la “decisión” de adjudicar estados mentales al sujeto en base a su propia comprensión, dejando por fuera el problema de la *intencionalidad*. Según Searle, existen tres tipos de *intencionalidad*: la *intencionalidad real*, la *intencionalidad derivada* y la *intencionalidad metafórica*. La primera es la que tienen los seres humanos y quizá los animales, consiste en la capacidad de mentalizar o traer a la mente objetos que no están presentes, por ejemplo, en este momento pienso en la ciudad de Valdivia y en sus calles, pese a que estoy recluido en mi casa de Santiago escribiendo, además, yo propiamente tengo la intención de escribir cada palabra, y ésta sale directamente de mí. La segunda corresponde a la “intencionalidad” del dinero, los mapas o la mera caligrafía (entre otros), por ejemplo, las palabras escritas en este texto no son más que rayones con un espacio de fondo blanco, podrían no significar nada, de hecho, por si mismas no significan nada, solo toman un sentido y una “intención” en el momento que son leídas por alguien, y no cualquier persona, sino alguien que sea hispano parlante, pues un sujeto cuya lengua materna y única es el inglés o el chino, difícilmente podría entender lo que está escrito aquí. Finalmente, la *intencionalidad metafórica* refiere a cuando se asume por términos prácticos que un objeto posee una intención. Doy de ejemplo el caso del cajero automático, una persona podría pensar o decir que “el cajero tiene el propósito de expeler dinero” pero, realmente aquella máquina solo tiene la función de sacar dinero, más allá de la ejecución del programa no existe ningún tipo de intención real de parte de este, del mismo modo que la mesa no “desea” estar firme o que un trompo no “planea” girar. Pero el criterio internalista de Turing se libra por completo de esta problemática, el manejo de un criterio internalista tiene un fundamento epistémico y no ontológico, realmente Turing no quiso decir qué es la inteligencia artificial, sino más bien cómo se podría estar medianamente validado a creer que se está frente a una, pero si se llegara a errar, no significa que una máquina no pueda pensar, solo que hubo un error; y en tanto a la capacidad para discernir sobre la elección, nuevamente, que se haya fallado no significa que se haya hecho solo de manera metafórica, solo que se hizo mal. Para explicarme doy como ejemplo un tirador disparando a una diana, no porque no acierte justo al centro, el hombre no disparó, simplemente hubo un factor que falló, sin

embargo, el método fue llevado a cabo de la mejor manera posible. Del mismo modo al adjudicar estados mentales tanto en el día a día, como en situaciones anómalas, se activan los mecanismos que tiene el ser humano para esta labor y a un nivel inconsciente (al menos por lo general) toma la decisión sobre si el otro tiene o no estados mentales.

1) El funcionalismo de Turing:

Para lograr un entendimiento más acabado de la filosofía de la mente de Turing, es necesario echar un vistazo a su concepción sobre qué es la mente humana. El profesor Rodrigo Gonzales de la Universidad de Chile clasifica a Alan Turing como un filósofo funcionalista, quienes consideran que la mente humana no se define por sus cualidades materiales, sino que por su función. Una definición funcional es por ejemplo la de una silla, esta no importa si está hecha de madera o metal, si tiene cuatro, tres o cinco patas, sino que sea útil a la hora de sentarse. En su texto *Descartes y Turing: dos criterios para la presencia de estados mentales* (González, Dos criterios para la presencia de estados mentales: Descartes y Turing, 2016) González lo expresa así¹⁰:

¿Por qué Turing introduce el tema del sexo de los participantes y su identificación?

Usualmente se ha desechado la identificación del sexo de los participantes por ser un elemento poco importante en el Juego de la Imitación. El carácter críptico de la primera versión del Juego podría haber ayudado a desestimar la importancia de dicha identificación. No obstante, parece razonable suponer que lo que Turing propone es una concepción funcionalista de los estados mentales, y ello se relaciona justa, aunque no exclusivamente, con la primera versión de su Juego. Lo dice explícitamente, pese a las acusaciones que se le han formulado de haber planteado un test puramente conductista. Según esta interpretación del Juego, lo único que importaría es el desempeño lingüístico del hombre dentro de la pieza A. No obstante, parece plausible creer que el funcionalismo de Turing refiere a la imitación de una capacidad, la lingüística femenina, reflejada mediante conducta. (González, Dos criterios para la presencia de estados mentales: Descartes y Turing, 2016)

¹⁰ La analogía entre la capacidad del hombre y la máquina de imitar a la mujer, representa el funcionalismo de Turing, ya que el parece creer en la plausibilidad de que el hombre y el robot lo hagan por igual, la instanciación material es poco relevante.

Parece evidente que la instanciación material no cumple otro rol más allá de ser eso, la instanciación material necesaria en la medida en que no se aceptan entidades reales no materiales, el hombre puede imitar mediante sus respuestas el mismo funcionamiento que debería tener el pensamiento femenino, del mismo modo se podría asumir que el computador digital sería capaz de imitar el funcionamiento de un cerebro humano, así el juez al evaluar las respuestas que den los participantes podría llegar a confundirse, si en realidad fuera necesario para haber estados mentales una instanciación material específica, el juez no podría confundirse, puesto que para el computador sería imposible *a priori* poder imitar a un ser humano.

2) La trampa lingüística

Turing genera todo el juego de la imitación a partir de una desconfianza personal sobre los términos en los que se habla, en su época “pensar” e “inteligencia” se adjudicaban generalmente a los seres humanos, mientras que el término “máquina” solía asociarse a una concepción mecanicista en que todo era movido en base a reacciones físicas, por lo cual hablar de una máquina que piensa era motivo de desconfianza “a-priori” (en un sentido amplio); a pesar de que luego repasaré las defensas que hace éste sobre la inteligencia artificial, es oportuno dejar en claro la intención que tiene (Turing) con este proyecto. Él parte su texto *Máquina computacional e inteligencia* (1950) de la siguiente manera:

Propongo considerar la siguiente pregunta: “¿Pueden pensar las máquinas?”. Se debiera comenzar definiendo el significado de los términos ‘máquina’ y ‘pensar’. Estas definiciones deberían ser elaboradas de manera tal que reflejen lo mejor posible el uso normal de estas palabras, pero una actitud así es peligrosa. Si el significado de las palabras ‘máquina’ y ‘pensar’ proviene del escrutinio de cómo son usadas comúnmente, se hace difícil escapar de la conclusión de que el significado y respuesta a la pregunta “¿pueden las máquinas pensar?” debiera ser buscado en una encuesta estadística, tal como la encuesta Gallup. Pero eso es absurdo. En vez de intentar una definición así, propondré reemplazar esa pregunta por otra, la cual se encuentra estrechamente relacionada y que se puede expresar en palabras relativamente poco ambiguas. La nueva forma del problema puede ser descrita en términos de un juego, el cual llamaremos “el juego de la imitación”. (Turing, 1950)

Debido a la “trampa lingüística” el propio Turing propone que ante tal temática la evasión de los problemas es una prioridad, ya que el proponer una definición sobre inteligencia o

pensamiento podría hacerlo caer en un error, sin embargo, parece ser que el pensamiento de Turing es contradictorio, en el texto anteriormente citado del autor, él junto con una descripción del funcionamiento de los computadores digitales muestra una clara inclinación a creer que la mente humana tiene el mismo tipo de funcionamiento que estos, teniendo un “input”, un conjunto de reglas y luego extrayendo un “output”, esto a manera de defensa de su teoría; si es que la mente humana funciona del mismo modo que lo hacen los computadores digitales, es posible fabricar uno que tenga estados mentales similares a los humanos. Podría argumentarse en consecuencia que Turing dio una definición de intelecto, sin embargo, el profesor González tiene argumentos suficientes para dilucidar que esta postura es realmente un mito, en su texto *El test de Turing: dos mitos un dogma* (González, *El Test de Turing: Dos mitos, un dogma*, 2007) él defiende que es una “ficción” refiriéndose a la siguiente cita sacada de una entrevista al mismo Alan Turing:

No quiero dar una definición de qué es pensar, pero si tuviese que darla, probablemente sería incapaz de expresar nada más acerca de ésta que decir que fue un tipo de zumbido mental [buzzing] en mi cabeza. Pero no veo que tengamos que estar de acuerdo en una definición en modo alguno. Lo relevante es tratar de distinguir entre las propiedades de un cerebro o las de un hombre, que queremos discutir, y aquellas que no queremos. Para ponernos en un caso extremo, no estamos interesados en el hecho de que el cerebro tenga la consistencia de la papilla. No queremos decir ‘esta máquina es muy compleja, luego no es un cerebro y no puede pensar’. Me gustaría sugerir una clase particular de test que uno pudiese aplicarle a una máquina. Ud. podría querer llamar a este un test para ver si la máquina piensa, pero sería mucho mejor no formularlo así y caer en la petición de principio, diciendo que las máquinas que pasen el test serán, por decirlo de algún modo, máquinas grado A. La idea del test es que la máquina tiene que simular ser un humano a través de responder a las preguntas que se le hacen, y sólo pasará este test si la simulación es suficientemente convincente. (Turing 1952, énfasis mío). (González 2007).

Parece en realidad que la intención de Turing al demostrar un cierto apego por el computacionalismo, fue meramente de un carácter retórico y nada más. En conclusión, se puede sostener que Turing realmente no tiene un compromiso conceptual al proponer el test, en cambio el compromiso sostenido es meramente metódico a favor de una posibilidad para acertar en el juicio sobre los estados mentales de la máquina.

3) La defensa de la IA y la importancia de Turing

Turing en el texto *maquinaria computacional e inteligencia artificial* dedica una gran parte a defender la posibilidad de la existencia de las IA el filósofo inglés parte con una predicción, a saber, que dentro de 50 años desde el texto (1950) habrá computadores capaces de ser programados para jugar, y que el lenguaje corriente ya habrá aceptado lo suficiente una idea como una máquina pensante. Se debe por fuerza reconocer que la ciencia ficción y el cine ayudaron a Turing en su segunda predicción lo suficiente como para que la idea de una IA no suene a contradicción de buenas a primeras, pero en cuanto a la primera predicción demostró lamentablemente que el Test de Turing es insuficiente, ya que a pesar de que el chatbot Eugene en el año 2014 paso el test, el hecho de si tiene estados mentales o no quedó inconcluso.

Turing detecta 9 posibles argumentos en contra de la inteligencia artificial, los cuales listados en orden son:

- i) La objeción teológica.
- ii) La objeción de las cabezas en la arena.
- iii) La objeción matemática.
- iv) El argumento desde la conciencia.
- v) Argumentos desde las discapacidades múltiples.
- vi) La objeción de lady Lovelace.
- vii) Argumento desde la continuidad del sistema nervioso.
- viii) Argumento de la informalidad de la conducta.
- ix) Argumento de la percepción extra-sensorial.

De entre los nueve argumentos solo repasare cuatro de ellos, debido a que los restantes son irrelevantes para este texto en específico.

- i) La objeción teológica: el argumento está en un diálogo directo con el dualismo de substancias, en resumidos términos, dicta que desde una visión teológica el alma y el intelecto son lo mismo y dios confirió alma solo a los hombres. Turing contesta desde una óptica teológica que se podría interpretar que el hombre solo crea un cuerpo para un alma creada por Dios, sin embargo, él prontamente

descarta el argumento diciendo que las discusiones teológicas han perdido su relevancia en vista de los avances que ha dado la ciencia¹¹.

- ii) Argumento de la conciencia: consiste en pensar que hasta que una máquina no sea capaz de tener emociones y escribir en base a esas emociones, simplemente no se la puede considerar como pensante, a lo que Turing responde que el juego mismo de la imitación puede dar señales sobre si la máquina es capaz de escribir un soneto o no, y luego responder preguntas sobre el mismo argumentando sus decisiones, en último término solo es exigirle más al computador participante.
- iii) El argumento de las discapacidades múltiples: este argumento dice en términos muy simples que es imposible hacer una máquina que pueda replicar todo lo que es capaz de hacer el hombre. Para empezar, es una versión diferente del argumento cartesiano que asumía que las máquinas son limitadas por su construcción, pero la defensa de Turing resulta bastante simple, y es que hasta ahora no se ha logrado construir una con una capacidad de memoria suficiente para imitar por completo al ser humano, sin embargo, dentro de las múltiples cosas que el argumento asume que una máquina no puede hacer, está la capacidad de equivocarse; este último punto es fácilmente salvable y es que se puede programar a la máquina para fallar de vez en cuando a una respuesta engañando a los jueces, en tal caso para el criterio internalista de Turing no importaría la regularidad que esta falla tenga.
- iv) La objeción de Lovelace: el argumento de Lovelace dicta que una máquina realmente no puede hacer nada nuevo, pero si se piensa bien esto no separa a la máquina del hombre, ya que muchos seres humanos en su vida no llegan a hacer nada realmente nuevo (que no haya hecho otro). Otra parte del argumento dice que una máquina no puede llegar a sorprender a una persona, Turing responde a esto último con un argumento internalista y si bien es mucho más refinado de lo que expongo aquí, la sorpresa en realidad se lo lleva quien es sorprendido y no necesariamente quien sorprende, de una manera más sencilla de entender doy el

¹¹ Realizar una objeción frente a la visión que pueda asumir la teología es necesario en la medida en que esta tiene como presupuesto que solo los seres humanos piensan en la medida que están dotados de alma y esta es la pensante, y una vía que defienda la emocionalidad tal como la mía requiere necesariamente del cuerpo del sujeto

ejemplo de una pareja, donde uno de los dos le roba un beso al otro mientras ven televisión, si bien no es la primera ocasión en que se besan, y probablemente ya lo han hecho en el mismo lugar, el sujeto besado se sorprende más porque él no se lo esperaba que por lo revolucionario del acto, por tanto, una máquina podría hacer algo común y corriente, pero si el sujeto que la observa no se lo esperaba se verá sorprendido, y en base al ejemplo anterior y el criterio internalista de Turing se puede interpretar que la máquina no necesitaría de creatividad.

Las tres defensas posteriores al argumento teológico realizadas por Turing que he expuesto, dejan escapar la debilidad de su propuesta, ya que esta resulta especialmente débil frente al ámbito emocional, si bien parece ser que Turing considera de manera muy lejana¹² la emocionalidad. Coherente con su sistema internalista, la defensa apunta solo a la capacidad de las máquinas para imitar al ser humano, y no para experimentar las emociones, ya que el internalismo de Turing en específico solo apunta a cualidades epistémicas y no metafísicas (¿cómo se puede asumir X? y no ¿qué es X?), sin embargo, sin estas defensas me sería imposible llevar a cabo el experimento mental que presentare en el capítulo cuatro y defenderé en el quinto.

Finalmente, frente al criterio internalista que ha llevado Descartes, y respetado por Turing, he decidido conservarlo. Ante el hecho ineludible de que el único conocimiento directo al cual se puede acceder es la propia mente, creo que cabe espacio para la duda en cualquier otra teoría imaginable, por lo cual me cuesta desprenderme de él, por tanto, al igual que estos dos filósofos, mi propuesta es de un carácter internalista, no sin tener diferencias en mi posición epistémica con ambos en la medida que contra Descartes no requiero de certeza, sino de convencimiento y a Turing en la medida que mi propuesta si requiere de brindarle atributos a la máquina en vez de solo crear un sistema mímico.

¹² Sobre todo, en la objeción de la conciencia él acepta que las maquinas deben imitar también la emocionalidad humana.

Capítulo 3: El debate actual, la ciencia cognitiva y teoría de la mente ToM

La ciencia cognitiva

La ciencia cognitiva es el nombre que se le da al trabajo interdisciplinario entre la *psicología cognitiva*, la *filosofía de la mente*, la *lingüística*, la *antropología*, la *neurociencia* y los estudios sobre *inteligencia artificial* dedicados al estudio de la mente humana con estándares científicos, teniendo a la recolección de datos y la contrastación empírica como paradigma principal.

Esta corriente tiene sus orígenes en el siglo XX como una alternativa al *conductismo*, que fue una corriente en psicología caracterizada principalmente por una visión científica que reducía los estudios psicológicos al estudio de la conducta animal y humana, con la fórmula estímulo (input) y respuesta conductual (output), fundamentados ante la creencia de que no podemos tener experiencia sobre los estados mentales ajenos, según C. Beorlegui, se puede clasificar a los conductistas de maneras a) conductismo *epistémico* versus el conductismo *ontológico* y b) conductismo *científico* contra el conductismo *lógico*. A continuación, daré una descripción breve de cada una de las subdivisiones de esta corriente (Beorlegui, 2007):

A.1 *conductismo epistémico*: parte de ignorar la existencia de lo mental (sin negar en principio que exista)

A.2 *conductismo ontológico*: a diferencia del anterior esta niega en principio la existencia de lo mental, dejando como único componente de lo humano la conducta.

B.1 *conductismo científico*: esta versión se centra en el estudio científico de la conducta para encontrar las leyes que la rigen, es la línea desarrollada por Watson, Skinner y sus discípulos. Asume que no existen diferencias cualitativas tanto en la conducta animal como la humana y gracias a la influencia de Pavlov, tienen una fuerte tendencia ambientalista, por lo cual la conducta u output depende en gran medida de las respuestas que haya recibido el sujeto de estudio desde la temprana infancia de parte del medio ambiente.

B.2 *conductismo lógico*: el conductismo lógico es una postura que intenta expresar los enunciados mentalistas con enunciados más simples, por ejemplo, si se dice que “Constanza cree que lloverá”, podemos reducirlo a “Constanza ve que el cielo esta nublado”, “Constanza toma el paraguas”, “Constanza se abriga”, etcétera, de este modo se permite evitar un lenguaje mentalista (y según esta corriente) no científico. El gran problema que tiene esta postura, o por lo menos según la visión de Searle, es su clara contradicción, pues el que “Constanza toma el paraguas” implicaría que “Constanza no desea mojarse”, y el desear es en sí mismo un estado mental, por tanto, el tomar un estado mental para explicar la conducta sin estados mentales es contradictorio, sin mencionar que es evidente que uno mismo tiene estados mentales. El problema es epistémico, no porque no exista la posibilidad de estudiar de modo científico los estados mentales de un individuo, significa que estos no existan o no tengan relevancia en la conducta.

Por otro lado, el surgimiento de las ciencias cognitivas consiste en un giro de vuelta al mentalismo, si bien ya no se toma desde una óptica dualista como la cartesiana, la existencia de estados mentales es el factor principal a estudiar en la medida que los cognitivistas comprenden a estos como causantes de la conducta humana. El cognitivismo tuvo una fuerte influencia de los desarrollos de los estudios sobre la *inteligencia artificial*, y a pesar de que una IA no ha sido creada aún (al menos no en los estándares que se buscan), se suele pensar que el desarrollo de la mente humana no debe diferenciarse mucho del de un computador digital. Entre las corrientes más destacadas se pueden encontrar:

- 1) *Teoría de la identidad*: esta postura se basa en la creencia de que mente y cerebro son equivalentes, del mismo modo que podemos considerar agua = H₂O. Para los teóricos de la identidad, un estado cerebral específico es equivalente a un estado mental, por ejemplo, el dolor es equivalente a la estimulación de las *fibras c*, el problema de esta postura radica principalmente en el principio de equivalencia según Leibniz, el principio de equivalencia consiste básicamente en que si A= B, todas las características de A también se cumplen en B; no es claro que el dolor sea igual a una estimulación de las *fibras c*. Otra corriente dentro de la teoría de la identidad que tiene un carácter más funcionalista es la versión que propone los *tokens* que funcionan como “instancias” específicas, de tal modo que a un estado mental específico le

corresponde una “instanciación específica, pero esto es aplicable a un solo organismo o especie, por ejemplo, la letra equis de la “fuente” “Times New Román” es “X”, mientras que para “calibri” “X” en ambos casos la manera en que se “instancia” o dibuja la letra es diferente e identificable con su fuente, pero el significante es el mismo, de igual modo un computador podría procesar “dolor” como la activación de un “chip” específico tras presionar un botón, y un ser humano con la activación de las *fibras c*, sin embargo no es claro que esta versión logre escapar de la ley de equivalencia de Leibniz

- 2) *Funcionalismo*: esta postura ya ha sido mencionada con anterioridad en el capítulo dos, sin embargo, es necesario repasar un poco más esta postura. Según esta corriente, como ya se ha dicho, la mente se define en tanto cumple una función y no de una manera “ontológica”, de tal forma, la comparación con un computador se vuelve inevitable, la instanciación material de la mente es un *hardware* mientras que la mente misma es un *software* o programa que se encarga de llevar a cabo las distintas operaciones. Julio Rodríguez lo expresa de la siguiente manera:

Se concibe la mente literalmente como una computadora. Los procesos cognitivos en la mente son equivalentes o aproximadamente equivalentes a una computadora. La Inteligencia Artificial (IA) concibe una computadora como un sistema formal automático: es decir, un conjunto de procedimientos de manipulación de símbolos regido por reglas lógicas, (Haugeland 1985) una secuencia finita de pasos (operaciones) para alcanzar el objetivo para el cual el procedimiento es diseñado (Johnson-Laird 1983, 1988), y allí donde el procedimiento tiene como expresión formal un algoritmo o programa.

Pero esta no deja de ser una visión muy pobre sobre el funcionalismo. Los funcionalistas en realidad estiman que la mente es un sistema complejo e interrelacionado que explica la conducta.

Arquitectura de la mente

La arquitectura mental es un concepto acuñado dentro del mismo marco de la ciencia cognitiva, resulta ser de vital importancia para el entendimiento de esta, debido a que en último término la plausibilidad de los estudios dependerá de haber acertado en esta. En palabras de Nicolás Medina:

Un aspecto importante en la ciencia de la cognición es la teoría de la arquitectura cognitiva, por medio de la cual se describe el marco estructural mediante el cual se realiza el procesamiento de la información, y consecuentemente sería responsable de las restricciones fundamentales de la conducta del organismo (o sistema). Específicamente, la arquitectura cognitiva se refiere al diseño y organización de la mente, que determina que el sistema cognitivo posea la característica de tener una estructura y una función. (Medina, 2008)

Ahora bien, la arquitectura mental no acaba las posibles variaciones que tenga la mente de un individuo, sino más bien, tiene el propósito de generar un cierto marco referencial en el cual puede ser procesada la información. Para una gran parte de los científicos cognitivos, la mente tiene un carácter “representacionalista” esto quiere decir que los estados mentales pueden ser equiparados a símbolos y estos símbolos a su vez computados, por lo cual la arquitectura mental sería el modo de computar los símbolos y no los símbolos mismos. Doy a modo de ejemplo una calculadora, si esta se le da de input 2X2 o 4X4 los símbolos serán distintos, pero el proceso de computo o algoritmo¹³ será el mismo. Esto por lo menos hasta la década del 90 donde aparecieron varias teorías alternativas, que en vez de trabajar con representaciones en cambio preferían modelos en los cuales el cuerpo y la experiencia tengan un rol más protagónico.

Teoría de la mente (ToM)

La *teoría de la mente*, o por sus siglas en inglés ToM (Theory of Mind) es una capacidad cognitiva que poseen los seres humanos, sin embargo, este término tiene una “carga” que lo inclina a los teóricos modulares, por lo cual prefiero utilizar el término “mindreading” (lectura de mentes) para denotar esta habilidad específica. La cual les permite ser capaces de adjudicar estados mentales a otros y a ellos mismos, y tema central de esta tesis. Según la ciencia cognitiva es un sistema de conocimiento que nos permite inferir que los otros son poseedores de creencias, deseos, sentimientos, y de esta manera conseguir interpretar, explicar o comprender los comportamientos propios y de otros, así como predecirlos y controlarlos. La cuestión es no menor, ya que, si no fuese por esta capacidad, los seres

¹³ Entiéndase algoritmo como una secuencia de pasos finitos con uno recursivo para llegar a un resultado, cabe destacar que ninguno de estos pasos requiere de tener una conciencia.

humanos serían incapaces de llevar relaciones y nexos sociales, lo cual los habría llevado probablemente a la extinción.

Los primeros atisbos de una “teoría de la mente” se le suelen adjudicar a Dennett, quien propone la “estrategia intencional”¹⁴ la cual postula que si se dota de creencias e intenciones a otro ser, es posible predecir su comportamiento, sin embargo, el término fue acuñado por Premack y Woodruff en su trabajo “Does the chimpanzee have a theory of mind” [Tiene el chimpancé una teoría de la mente] (1978) (Martinez, 2011)

En un experimento, Premack y Woodruff sometieron a un chimpancé a videos en los cuales se mostraba a un hombre enfrentando distintos problemas, ya sea alcanzar unos plátanos o el mismo sujeto encerrado en una jaula, a continuación, se le mostraba una serie de imágenes en las cuales una tenía la solución, y sorprendentemente solía acertar en la imagen, lo cual demostraba que los chimpancés reconocían la existencia de un problema y las intenciones del sujeto, sin embargo, no fue definitivo si es que eran capaces de adjudicar estados mentales (Zilber, 2017) luego los mismos científicos fueron capaces de hacer pruebas más allá, demostrando que los animales eran capaces de engañar a un humano escondiendo su alimento; si bien la prueba del engaño fue considerada como fundamental en los simios, para los niños se tomó como paradigma principal la prueba de la falsa creencia, que será expuesta en la siguiente figura:

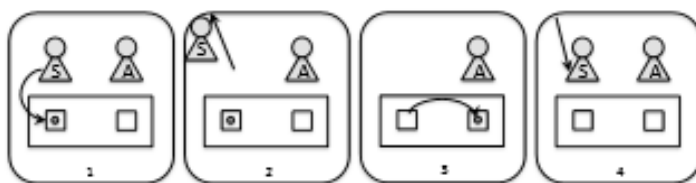


Fig.4

Un niño (A) que es sometido a la prueba es sentado en una mesa, junto a él otra persona que en la figura es tipificada como “S” deja una galleta en un lugar que no se ve a simple vista (paso 1), luego “S” se retira del lugar (paso 2) y un agente externo mueve la galleta de lugar

¹⁴ Esta es sólo una estrategia epistémica, no hay compromiso ontológico con la existencia ni siquiera de mentes. Aquí la mente del otro sólo está en el ojo del observador.

a otro en que no se vea a simple vista (paso 3) y a la vuelta de “S” se le pregunta a “A” en qué lugar cree “S” que estará la galleta. Los resultados dieron lugar a que la gran mayoría de los niños menores a cuatro años cometieran un error egocéntrico y dijeran que “S” buscaría la galleta en la segunda ubicación (en otras palabras, no fueron capaces de diferenciar la mente de “S” de la suya), en cambio los niños mayores de cinco años, fueron capaces de detectar que “S” buscaría su galleta en el lugar que la dejó. La experimentación muestra que los niños a partir de un rango de edad entre los tres y los cinco años desarrollan una ToM.

Sin embargo, ToM no es una cuestión de todo o nada, el experimento no implica que los niños a partir de los 5 años ya tengan una ToM “madura”, en cambio es muy debatido si es un proceso que va evolucionando a lo largo del tiempo y como es que lo hace¹⁵.

Frente al problema de cómo surge y se desarrolla la ToM, existen varias posturas entre ellas, las dos más importantes son la *teoría de teorías* y la *teoría de la simulación*. Sin embargo, en el último tiempo han surgido nuevas teorías que han sido llamadas “alternativas” entre las cuales se ubica el propósito de esta tesis, pero serán explicadas en los capítulos siguientes. A continuación, presentaré la *teoría de teorías* y la *teoría de la simulación*.

1) Teoría de teorías

Una buena manera de abordar el enfoque de “teoría de teorías” (TT) es mediante la metáfora del “niño científico”¹⁶ (Delgado, Los niños y la comprensión del mundo mental de los otros: Teorías explicativas, 2018) esta consta de considerar al niño como nacido con una cierta teoría sobre cómo funcionan las mentes humanas, incluidas la suya, y mediante la observación y predicción de la conducta ajena, este la va puliendo hasta que entra en una situación de “crisis de paradigma” y la intercambia con otra constituida por los argumentos *ad hoc* utilizados para salvar la teoría anterior.

¹⁵ Si bien esto es un debate aún vigente en este texto he decidido decantarme por esta versión, ya que considero que en caso de que los teóricos de tercera persona tengan razón esta sería la más probable.

¹⁶ Propuesta por A. Gopnik y H. Wellman en su artículo *Why the Child's Theory of Mind Really Is a Theory* (Gopnik & Wellman, 1992)

Esta visión tiene un carácter netamente racionalista, donde el individuo nace con “módulos”¹⁷ innatos (al menos en la versión de Baron-Cohen) (Zilber, 2017) que procesan la información (metarepresentaciones) de manera organizada, así el niño va adquiriendo a medida que se desarrolla, maneras más complejas de metarrepresentaciones para llegar a realizar representaciones simples sobre objetos en el mundo y representaciones más complejas sobre entidades mentales. El juego de la falsa creencia, según esto, se explicaría en que los niños menores de cuatro años solo son capaces de realizar representaciones situacionales, mientras que los niños mayores a los cinco años en cambio son aptos para realizar metarrepresentaciones.

La estructura original de Baron-Cohen (2000) consta de cuatro módulos, 1) detector de intencionalidad (ID), 2) detector de dirección ocular (EDD), 3) mecanismo de atención compartida (SAM) y 4) mecanismo de teoría de la mente (ToMM). El ID y el EDD aparecen en los primeros 9 meses de vida y construyen representaciones diádicas de estados mentales simples, volitivos (ID) y perceptuales (EDD); mientras que el ID atribuye deseos o metas a lo que interpreta como agentes, el EDD atribuye percepciones a partir de la dirección de la mirada. El SAM, en cambio, aparece entre los 9 y los 14 meses y construye representaciones triádicas, a partir de los inputs que recibe desde el ID y el EDD, interpretando si el “self” y otro agente están percibiendo el mismo evento o no. El ToMM, surge entre los 2 y los 4 años, con la función de representar un estado epistémico sobre los estados mentales, incluyendo las emociones (Zilber, 2017), culminando en que la mayoría de los niños puedan dar una respuesta satisfactoria en la prueba de la falsa creencia, sin embargo, se puede culpar a esta teoría de caer en una intelectualidad exacerbada; por mucha flexibilidad que tengan estos módulos para formar metarrepresentaciones, nunca pierde el carácter racionalista, apoyándose en el lenguaje (más bien la posesión de lenguaje, que el lenguaje mismo) para explicar su funcionamiento y en contraste la emocionalidad debe ser explicada teniendo en cuenta esta representaciones. No obstante, las maneras en que los módulos debiesen operar sería confusa, las emociones presentan un rol causal frente a las representaciones, de tal modo que las representaciones que provea el programa debiese ser influido por el módulo de

¹⁷ Un módulo es una “unidad” mental que se dedica a procesar cierto tipo de metarrepresentación.

emocionalidad, pero al ser las emociones también representaciones (para el modelo de teoría de teorías) las mociones debiesen en algún momento ser causa de sí mismas.

2) Teoría de la simulación

A diferencia de la teoría de teorías, esta tiene un enfoque empirista y de primera persona (la teoría de teorías tiene un enfoque de tercera persona) si es que la anterior se explicaba mediante la metáfora del “niño científico” esta, en cambio, puede ser explicada mediante la metáfora de “el teatro de la mente” que en resumidos términos expresa que la mente del sujeto pensante “simula” ser el otro e intenta predecir la conducta ajena en base a la pregunta ¿Qué haría yo en su lugar?

Esta postura nace directamente de un rechazo hacia la postura de teoría de teorías, en la cual el enfoque intelectualista resulta predominante. Es posible remarcar dos versiones predominantes de la teoría de la simulación, a) la introspección y b) la proyección (Gordon 1992)¹⁸ en el caso de la proyección, es el yo del individuo quien se proyecta en las experiencias del otro, o en otras palabras (“si yo fuese quien ha vivido lo que él vivió ¿Qué haría?, ¿Qué pensaría?”) en cambio, la introspección implica tener que fingirlo. (“¿si fuese **yo**, que haría **yo** en su lugar?” (Delgado, Tres debates fundamentales en el campo de la teoría de la mente: Aspectos teóricos y metodológicos, 2011).

Siguiendo el hilo explicativo de Delgado, es necesario abordar la posición de sobre la teoría, él consideró que los niños no necesitan construir una teoría para comprender la mente de los otros, en oposición a ello, este autor defiende que la comprensión de la mente de los otros se basa en mecanismos intuitivos, con una clara tendencia fenomenológica. Este conocimiento intuitivo está inundado de elementos autorreferenciales, según Johnson, la intuición mentalista de los otros es en realidad una referencia a él mismo, en este sentido, la capacidad de proyectar e imaginar la propia experiencia en otros, es el mecanismo utilizado por el niño para comprender los estados mentales de los otros (J. Delgado 2018) así, a través de la propia proyección de las experiencias internas, es que entiende las acciones de otros. Una metáfora que suelen dar los teóricos de la simulación es la del “avión en la turbulencia” si es que

¹⁸ (Delgado, Los niños y la comprensión del mundo mental de los otros: Teorías explicativas, 2018)

alguien quisiese saber cómo se comportaría un avión frente a una ráfaga de viento, le convendría más realizar un modelo del avión y someterlo a una cámara de prueba, antes que recolectar todos los datos y contrastarlos con fórmulas matemáticas; la metáfora sirve como argumento a favor de la perspectiva de primera persona, en la medida que es preferible usar la propia mente de un sujeto a modo de cámara de simulación, en vez de constructos teóricos, Esto queda más claro al revisar la explicación que realiza Analía Zilber sobre la teoría de Harris (Zilber, 2017) , según la autora, Harris se posiciona desde la perspectiva de introspección, donde el niño mediante la información privilegiada que tiene de sus propios estados mentales, imagina en este “juego” los estados que debe tener el otro, de este modo el infante desecha cualquier constructo teórico dejando solo la interpretación de su “teatro interno” (Zilber, 2017), pero esto no significa que esta capacidad imaginativa nace completamente desarrollada, por el contrario, es una cualidad que va evolucionando con el paso del tiempo, siendo dividida por Harris en cuatro etapas:

- 1) La primera etapa se da a partir del primer año de edad, los niños solo pueden “procesar” las emociones y voliciones que otros individuos junto a ellos tienen sobre aquellos objetos presentes en el entorno inmediato, pero esta “lectura” sobre los otros es on-line, y eso explicaría por qué los pre-infantes son capaces de mirar en la misma dirección o imitar los gestos de los padres.
- 2) En la segunda etapa, durante el segundo año de vida, los niños poseen una capacidad simuladora on-line, lo cual significa que pueden simular las percepciones y los sentimientos de las otras personas, lo que les permite atribuir actitudes intencionales hacia objetos del contexto presente y en consecuencia intervenir sobre ellas para regular la conducta de los otros con respecto a esos objetos. Esto explica que en esta etapa los niños puedan redirigir la mirada de los otros señalando o mostrando un objeto de interés, o también buscar modificar la actitud emocional de los otros sacándoles u ofreciéndoles un objeto deseado.
- 3) En la tercera etapa, al finalizar el segundo año de vida, los niños poseen una capacidad simuladora off-line, es decir, la simulación se desliga del contexto presente, lo que les permite imaginar actitudes intencionales de los otros, diferentes a las propias. Esto explica que en esta etapa los niños puedan anticipar o “representar” las reacciones de

personas cuyos estados mentales difieren de los propios (por ejemplo, cuando el otro ve o desea un objeto diferente).

- 4) En la cuarta etapa, alrededor de los 4 años y medio de vida, los niños pueden imaginar actitudes intencionales de otras personas respecto de objetos contrafácticos o hipotéticos, es decir, opuestos a los que ellos mismos perciben. Esto explica que en esta etapa los niños puedan reconocer que los otros pueden tener estados mentales que entran en conflicto con los propios (por ejemplo, cuando el otro ve o cree algo diferente con respecto a la misma situación), y por lo tanto comprender falsas creencias de primer orden.

De este modo los niños fallarían la prueba de la falsa creencia, no en tanto que no pueden simular la mente del otro, sino que el sistema de simulación que el niño posee aún no es lo suficientemente sofisticado, además Harris propone que existe una clara tendencia en los experimentos en favorecer a los teóricos de teorías, ya que en primer lugar las pruebas tienen un enfoque de tercera persona, por lo cual no dicen nada sobre los estados internos del niño; si ponemos el ejemplo de la galleta en el experimento de la falsa creencia, el niño “A” no buscaría la galleta en el mismo lugar que “S”, pues sus estados mentales no son los mismos.

Si bien el debate entre estas dos posturas aún no queda zanjado, nuevas se han adherido al campo de discusión, entre ellas la postura de Antonio Damasio, que será expuesta en el siguiente capítulo junto con la importancia de las emociones dentro de la teoría de la simulación y la teoría de teorías.

Capítulo 4: Sobre las emociones

Como ya se ha dicho en el capítulo dos, las personas rara vez se detienen a pensar sobre si el resto de los seres con los que se encuentran en su día a día tienen o no estados mentales, por lo general, los seres humanos asumen “a-priori” que los demás tienen estados mentales, pues la gran mayoría de la especie está dotada de “mindreading”, pero el objetivo de esta tesis es probar que la emocionalidad es un criterio cuando menos importante para la presencia de estados mentales, y a un nivel epistémico, que estas (las emociones) son un elemento

relevante para el “mindreading”, o en otras palabras probar que para el mindreading es primordial¹⁹ la emocionalidad frente a otros criterios (de una manera simple, es posible aseverar que otro tiene estados mentales si es que tiene emociones y se puede saber si es que tiene emociones mediante la propia emocionalidad). De todas maneras, para ambos objetivos es necesario brindar un caso de estudio principal, en el cual se enmarque la argumentación y se pueda recurrir como “caso anómalo”, sin embargo, esto no implica que este sea el único caso de estudio o ejemplo que vaya a tratar, pero no dejaré de referirme a este, y es a este caso al que pretendo dar solución finalmente.

He decidido tomar como ejemplo paradigmático el videoclip de la canción *I hope you're happy*²⁰ de la banda de rock norteamericana Blue October (Blue October, 2018). El motivo particular por el cual he decidido tomar este material audiovisual como ejemplo paradigmático es porque presenta un caso que resulta anómalo, y en este ejemplo en particular el mindreading de las personas parece atentar contra el sentido común. La utilidad del video radica en que “la” protagonista es un autómeta (siendo este el punto en el que se presenta la anomalía), y según los criterios presentados por Descartes, una persona al percatarse de que se trata de una máquina y no de una mujer debería dejar de adjudicarle estados mentales, sin embargo, el video está construido de tal manera que esto no ocurra. Hay que reconocer el hecho de que en realidad no se está frente a una máquina, sino, de una actriz, pero si la cuestión se funda en este hecho, el rol de la emocionalidad realmente no queda descartado, más bien se ve realzado, puesto que racionalmente no podemos concluir que no tiene estados mentales. Aun estando fuera del “mundo” del video, y analizando el caso dentro de todo uso de razón, la humanidad de la figura evoca la adjudicación de estados mentales, pese a que no se cuenta con evidencia que apoye el hecho de que posee mente más allá de ser una actriz. Sin embargo, hay que dejar sentado que el análisis sobre ¿Por qué es tan difícil dejar de adjudicarle estados mentales? desde las perspectivas de la teoría de teorías y la teoría de la simulación resulta mucho más complejo que este examen breve que he expuesto tomando distancia de una postura cartesiana, por lo cual, más adelante dentro de este capítulo será abordado este punto.

¹⁹ Entiéndase primordial en el sentido “antes que x”

²⁰ Pido al lector tomar un momento y revisar el videoclip cuyo enlace es el siguiente: <https://www.youtube.com/watch?v=FwZvuV79jeY>

El video además tiene la virtud de no tener dentro de la cinemática ningún dialogo verbal entre personajes, por lo cual los criterios lingüísticos no son relevantes para su análisis y se podría trabajar desde la emocionalidad y el lenguaje no verbal²¹. El cortometraje en cuestión lamentablemente se encuentra de manera anacrónica y consta de varias cinemáticas irrelevantes dificultando su análisis, por cual he procedido a ordenar y seleccionar de la mejor manera posible los momentos relevantes para la observación.

Descripción del metraje:

“Un hombre encuentra abandonado y en mal estado un autómata con apariencia femenina en un jardín. De inmediato decide recogerle y llevarlo a lo que se puede suponer es su casa, lugar en el cual procede a repararle, al terminar sus reparaciones lo activa y al notar la respuesta de la máquina que lo observa, casi cae al suelo de espaldas por la sorpresa. Posterior a eso el sujeto se dedica a enseñarle a bailar a la máquina, (cabe destacar que gracias a la actuación de la actriz parece carecer de expresiones faciales complejas). En un momento no determinado de la clase de baile dan a entender que se besan, no obstante, hay un corte de escena que impide ver el momento. El video da un salto y muestra al sujeto llevando al autómata a una fiesta familiar donde son recibidos en la puerta del lugar por una mujer mayor que realiza un gesto de extrañeza al ver llegar a la pareja, luego se los ve en un patio donde están el resto de los invitados, siendo ahora recibidos por un hombre mayor. Ya en la fiesta la pareja (el hombre y el autómata) proceden a bailar, instancia en que se besan (se puede interpretar como un segundo beso). Sin darnos un orden cronológico respecto al resto del metraje nos muestran una imagen en que el autómata observa un ave en una jaula (es válido suponer que es en la fiesta porque lleva la misma ropa que en las otras escenas) luego prosiguen a mostrarnos que la supuesta mujer juega con una piñata golpeándola en reiteradas ocasiones contra el piso, hasta que el bate que utilizaba se rompe, de tal modo que un trozo le atraviesa la palma de la mano. Ante esto el autómata mira su mano abriendo los ojos con una expresión que asemeja a la sorpresa, seguido de esto la robotiza es mojada, liberando un haz de luz por los ojos, provocando la sorpresa y miedo de los espectadores. El hombre la saca del patio ante lo ocurrido para luego recibir otro salto cronológico del video tras el cual

²¹ A pesar de que, en este punto, doy por sentado que existe una relación estrecha entre el lenguaje no verbal y la emocionalidad, también daré una explicación sobre el motivo por el cual trabajo con este presupuesto.

llega un hombre a la puerta del lugar. El recién llegado procede aparentemente a quitarle un “chip” al autómata (cosa que parece reiniciarla) y llevársela, entre tanto la suben a un vehículo, la robotiza procede a girarse y manifestar una sonrisa por primera vez, o por lo menos primera vez que lo hace de esa manera en cámara”.

El video presenta varias dificultades para el análisis, entre las más importantes está la que ya he nombrado (el hecho de tratarse de una actriz y no una máquina real), en segundo lugar, se trata de una manifestación artística y no un experimento científico, esto implica que controlar las posibles variables que se presenten al análisis, es cuando menos complejo, sin embargo, esto tiene otra virtud, ya que al precio de sacrificar precisión en cambio se recibe no estar sesgado por una postura, con esto me refiero a que la evidencia no tendrá como objetivo principal apoyar a alguna postura específica (haciendo preguntas sesgadas por ejemplo) y al no poder aislar las variantes del caso, será más cercano a la experiencia cotidiana. Haciendo un contraste con el experimento de la falsa creencia, este tiene el defecto de situar al niño en una posición en que debe hacer un juicio, contestando en base a ese juicio, y en la medida en que se suele premiar a un niño más por el acierto que el error, este podría dar la respuesta “correcta” sobre donde está la ubicación de la galleta, ignorando los estados mentales del otro, en este sentido el error narcisista no sería porque el niño no le adjudica estados mentales al otro (quien busca la galleta), sino que por simple narcisismo de tener la razón²²²³. Otra dificultad que presenta el video es la presencia de música, la cual puede estimular la emocionalidad de los espectadores, pero por lo menos el resto del capítulo, la historia del video será tratada exclusivamente como un experimento mental²⁴.

Durante la exposición del video, intenté de la mejor manera posible evitar hablar del autómata como si tuviese estados mentales, aun así, existen momentos que pueden provocar dudas, siendo el más importante aquel en que el autómata sonríe; una sonrisa suele ser considerada como un “output” conductual de un estado mental, generalmente siendo este la felicidad. Se puede mencionar también que los seres humanos pueden sonreír frente a alguna situación

²² Con esto último no quiero dar a entender que la prueba de la falsa creencia esté mal planteado o mal llevado a cabo, solo doy un ejemplo útil para el propósito de este texto, sin tomarlo como cierto.

²³ En el caso descrito, no es que el niño no entienda que el que busca la galleta esté en una falsa creencia, sino que el niño no entiende la intención del interrogador.

²⁴ Por cuestiones de tiempo y habilidades en esta tesis no he podido eliminar esta variable con algún experimento científico, sin embargo, en algún futuro trabajo pienso tratar este problema.

que les resulte incomoda o fingirla para obtener algún fin particular, algunas de estas son voluntarias y otras no, por lo cual se puede pensar en una relación entre un estado mental y un output conductual en la medida en que los pensamientos que se tengan coincidan con las emociones, por ejemplo, la aceleración cardíaca que se produce en un individuo al encontrarse con la persona de la que está enamorado (pura emoción); aun cuando el mismo sujeto tratase de fingir (estado mental) que se encuentra enamorado de otra persona sin tener aquellos sentimientos, no podría acelerar su ritmo cardíaco al ver a la víctima de su engaño utilizando solo la fuerza de mente y el control que esta tiene sobre el cuerpo, en cambio si el sujeto piensa en la persona de la que está enamorado, lo más probable es que experimente las respuestas emocionales en ese momento.

Para efectos del experimento mental que estoy tratando, este breve análisis que he realizado implica que es imposible decidir si el robot realmente tuvo sentimientos u emociones al momento de sonreír, al menos no sin construir algún modelo o arquitectura mental compleja que explique los procesamientos de la máquina, pero ese no es mi objetivo, y más allá de asumir que la máquina tiene múltiples circuitos que de algún modo imiten el comportamiento de un cerebro humano, el experimento no da certeza sobre este punto, en cambio, pretendo presentar otro camino. Según el análisis realizado, una posible causa para una sonrisa es el engaño, por cuanto para engañar es necesario tener estados mentales, en caso de que el robot quisiese engañar a su pareja de baile, necesariamente tendría que tener algún tipo de estado mental, no obstante, tras este “razonamiento” se encuentra una petición de principio, ya que asumo que tiene estados mentales (intención de engañar) para probar que hay estados mentales, así que la ruta debe ser por necesidad otra. Ahora consideraré que la sonrisa fue genuina, que realmente se reflejó un estado mental y más aún, que este fue involuntario, aunque de nueva cuenta se asume que tiene estados mentales, existe una diferencia con la voluntariedad del anterior; un pensamiento voluntario es claramente un estado mental, mientras que un pensamiento involuntario como las emociones no requiere realmente que exista una mente, ya que las emociones podrían confundirse con los cambios fisiológicos experimentados por el sujeto. Pero para entender este punto es necesario explicar qué es la emocionalidad y los sentimientos con la finalidad de luego exponer la respuesta involuntaria del autómata.

La distinción entre las emociones y los sentimientos

Una de las mayores dificultades para explicar la diferencia entre las emociones y los sentimientos se haya en una confusión lingüística presente por lo menos en el español coloquial: Los hablantes del español generalmente no hacen distinciones al hablar de sentimiento o emoción. Para explicarme doy el ejemplo de una emoción precisa, la “alegría”²⁵, cuando un hablante español manifiesta estar alegre dice “me siento alegre”, de igual modo se suele hablar de las emociones solo con el inicio del estado “anímico” por ejemplo “me he emocionado, me puse feliz”. Además, se tiene otra confusión lingüística, que se presenta en el verbo “sentir”, generalmente se habla de sentir independientemente si se trata de una sensación o un sentimiento, doy de ejemplo dos frases: “siento frío” y “me siento feliz”; la primera respecta a una sensación que depende de un sentido, que para efectos de este texto se tomará como la recepción de un input externo, mientras que la segunda frase respecta directamente a un estado mental, y durante lo que reste del texto intentaré aclarar a cuál de las dos acepciones me estoy refiriendo, por lo menos cuando considere que existe algún tipo de ambigüedad.

Volviendo a la confusión que existe entre el sentimiento (mental) y la emoción, esto se debe a que ambas tienen una estrecha relación. Por lo general se hace la distinción entre una y otra, por ejemplo, en el caso de la alegría y la felicidad con una graduación de intensidad, considerando a las que legítimamente son emociones como “menos intensas” que los sentimientos, la alegría es menos intensa que la felicidad o la atracción que el amor. Sin embargo, esta noción está errada, aunque no por completo sí en su substancia por lo menos, la distinción en realidad se encuentra en el lugar que se manifiestan. Por una parte, las emociones, de hecho, son fisiológicas, mientras que los sentimientos son puramente mentales, Antonio Damasio en una entrevista a Eduard Punset lo explica de la siguiente manera:

Es muy importante distinguir entre la fase de la emoción y la fase del sentimiento. Cuando experimentas una emoción, por ejemplo, la emoción de miedo, hay un estímulo que tiene la capacidad de desencadenar una

²⁵ Más adelante en el texto explico por qué la alegría es una emoción y no un sentimiento.

reacción automática. Y esta reacción, por supuesto, empieza en el cerebro, pero luego pasa a reflejarse en el cuerpo, ya sea en el cuerpo real o en nuestra simulación interna del cuerpo. Y entonces tenemos la posibilidad de proyectar esa reacción concreta con varias ideas que se relacionan con esas reacciones y con el objeto que ha causado la reacción. Cuando percibimos todo eso es cuando tenemos un sentimiento. (Damasio, El cerebro, teatro de las emociones, 2008)

Al menos desde la postura de Damasio las emociones tienen un rol causal frente a los sentimientos, la palabra emoción tiene su origen en el latín *mocion*, que a su vez significa *movimiento*, por lo cual podría interpretarse la emoción como una especie de “movimiento” que sufre el cuerpo, este movimiento es más bien metafórico, puesto que si bien hay partes específicas que se mueven durante la emoción, realmente se trata de un “estado” específico donde intervienen el funcionamiento de los órganos, y una reacción química específica del cerebro; en la emoción de atracción el corazón se acelera, una serie de procesos químicos internos generan una sensación de “cosquilleo en el estómago”, las pupilas se dilatan, la sangre fluye con más velocidad producto de la aceleración cardíaca concentrándose en lugares específicos como las mejillas, provocando que estas tomen una tonalidad rojiza, las manos sudan entre otras reacciones, sin embargo lo más interesante es que muchas afectan al comportamiento mental, algunas personas tartamudean (producto de que el nerviosismo le impide elegir bien sus palabras), haciéndoles difícil la comunicación, y otras toman decisiones que generalmente no harían como salir de su zona de confort para pasar más tiempo con el objeto del cual se encuentran enamorados, tal como cambiar las preferencias musicales o asistir a un evento que en realidad les desagrada o les parece poco atractivo. Damasio en el texto *En busca de Spinoza* (2005) presenta una definición de las emociones que resulta funcional para su análisis, y sin esta, la cita de la entrevista no puede ser entendida, pues es la que él tiene en mente al comentarla, la definición es la siguiente:

- 1. Una emoción propiamente dicha, como tristeza, vergüenza o simpatía, es un conjunto complejo de respuestas químicas y neuronales que forman un patrón distintivo.*
- 2. Las respuestas son producidas por el cerebro normal cuando éste detecta un estímulo emocionalmente competente (un EEC), esto es, el objeto o acontecimiento cuya presencia, real o en rememoración mental, desencadena la emoción. Las respuestas son automáticas.*

3. *El cerebro está preparado por la evolución para responder a determinados EEC con repertorios específicos de acción. Sin embargo, la lista de EEC no se halla confinada a los repertorios que prescribe la evolución. Incluye muchos otros aprendidos en toda una vida de experiencia.*

4. *El resultado inmediato de estas respuestas es un cambio temporal en el estado del propio cuerpo, y en el estado de las estructuras cerebrales que cartografían el cuerpo y sostienen el pensamiento.*

5. *El resultado último de las respuestas, directa o indirectamente, es situar al organismo en circunstancias propicias para la supervivencia y el bienestar (Damasio, En Busca de Spinoza, 2005)*

Además, se puede considerar que las emociones pueden ser calificadas como estados mentales. Es cierto que Damasio considera que las emociones son predominantemente fisiológicas, y su nexa con la mente humana consiste en ser la causa de los sentimientos y a la par, las emociones no son identificables con un pensamiento²⁶, sin embargo lo cierto es que una emoción se presenta en conjunto con una cierta “disposición” que es captada por la mente para generar un sentimiento (para Damasio un sentimiento se presenta cuando un sujeto toma constancia de un conjunto de emociones y estados corporales, a los cuales se le adjudica una cierta valoración de “agrado” o “desagrado”), esta “disposición” no es del mismo tipo en que se capta un dato sensorial, por ejemplo, escuchar una canción que al sujeto le recuerda a su ex pareja con la cual no quedó en buenos términos; en este ejemplo el oído del sujeto realizará las mismas operaciones que realizaría con cualquier sonido, el cerebro interpretará los sonidos y llegará a la conclusión de que el sonido particular es música, interpretará la letra dándole un significado y la memoria hará el nexa con su ex pareja, finalmente todo el aparataje químico desembocará en que se presente una emoción (para efectos de este experimento mental pueden ser la ira o la tristeza) y el sujeto tendrá pensamientos acordes con tal emoción, en el caso de experimentar ira podría recordar alguna discusión con la otra persona o uno de los malos hábitos de esta y en el caso de la tristeza, el momento del quiebre o algún momento preciado; es aquí donde se da un componente mental y se separa de las sensaciones. La información brindada por los sentidos generalmente depende más de factores externos, en el caso del oído se puede tomar en consideración: el

²⁶ En este contexto específico, con pensamiento me refiero a una representación mental, ya sea una imagen, un recuerdo o una situación inventada por el sujeto.

medio que emite el sonido, la distancia del oyente y el volumen del sonido, pero en contraste no existe necesidad entre el estímulo y la emoción específica, si bien lo normal es que se presente una emoción frente a la canción, el mismo sujeto podría tener distintas emociones (ira o tristeza en este caso) para el mismo estímulo, con esto me refiero a que la canción podría ser reproducida con los mismo parlantes, a igual distancia e igual volumen y aun así desencadenar la ira o la tristeza. Esto podría tener como explicación el funcionamiento del cuerpo, (en la medida en que las emociones se presentan necesariamente como un estado fisiológico) descartando el estímulo como causante de la variación. Ahora bien, se podría asumir que esto se debe solamente a las sustancias químicas que se encuentran presentes en el cuerpo al momento de recibir el estímulo, pero la naturaleza de las emociones no se presenta tan simplemente como cualquier otro proceso fisiológico, para ello contrastaré dos emociones distintas, el hambre y la tristeza. Por un lado, el hambre es la respuesta que tiene el cuerpo frente a la falta de alimento, siendo un “instinto” primitivo que le permite al ser vivo alertar que necesita conseguir energía y orienta sus deseos para obtener estos alimentos; en esta medida se puede diferenciar sin problemas que el cuerpo (para el hambre sería el estómago) está en una situación “X” y la mente realiza otro proceso totalmente distinto, “desear” para solucionar “X”. En cambio, en el caso de la tristeza, el cuerpo se encuentra en un estado “Y”, pero la mente es incapaz de localizar tal “sensación” en una parte del cuerpo, de tal forma la solución para el malestar fisiológico radica en adquirir algo para sí misma (la mente). Volviendo al caso de la ruptura amorosa y la tristeza, es posible interpretar que esa situación “Y” es el estado neuronal que provoca un malestar, pero no solo fisiológico, sino que también mental, realmente el cuerpo mismo no se encuentra bajo ninguna carencia, aún si se dijera que es quizá la necesidad de tener una pareja para la procreación, la emoción que se presentaría probablemente sería la excitación o una subida de la libido, o si fuese el deseo de protección, lo más lógico es que la emoción sea temor, y aun cuando puede que el sujeto tenga la emoción de temor, podría experimentar de igual modo la tristeza, por lo cual la “carencia” (si es que la emoción se presenta por carencia) sería netamente social, y las relaciones sociales se dan a un nivel mental y no fisiológico, por tanto es posible concebir que la emoción “tristeza” tiene una dimensión fisiológica en tanto se instancia materialmente con un estado específico del cuerpo y también tiene una dimensión mental. Con esto último no quiero decir o dar a entender que un ser humano puede experimentar una emoción

mentalmente primero, y que luego esta se presente “materialmente”, en cambio estoy bastante de acuerdo con Damasio en la necesidad de una “materialidad”, y que esta precede a la mentalidad de la emoción, en otras palabras, un EEC específico puede desencadenar una emoción bidimensional²⁷ en la medida que resulta casi imposible para el sujeto diferenciarlos, y la causa no tiene por qué ser una necesidad fisiológica.

La existencia de emociones bidimensionales trae consigo una complicación, y es que se deben graduar y clasificar las emociones con tal de saber cuáles son bidimensionales y cuáles simplemente tiene nexos causales con los estados mentales, la dificultad de esto radica en que todas las emociones conllevan un sentimiento consigo, y la diferencia entre unas y otras es muy leve en algunos casos. Damasio divide las emociones en tres clasificaciones, pero antes de repasar las clasificaciones dadas, es necesario hacer un breve repaso sobre el origen de las emociones en los seres vivos. El punto tres de la definición que utiliza Damasio hace una mención a la evolución. La evolución tal como se entiende hoy es un proceso azaroso, y lejos de funcionar como lo haría un arquitecto que planifica cada una de las partes de su obra, en cambio atiende a una serie de mutaciones genéticas que se dan en los seres vivos, siendo las que resultan beneficiosas las que se heredan de generación en generación, ya que le permiten al organismo adaptarse al medio ambiente el tiempo suficiente para reproducirse (por lo menos), en cambio las que impedían una adaptación óptima eran “desechadas”, ya que el organismo muere antes de pasar sus genes a la siguiente generación. Es así como los organismos que desarrollaron un mecanismo que les permitiese regular mejor su química interna tuvieron mayor oportunidad de reproducirse y llevar un mayor tiempo de sobrevivencia. Damasio deja claro que esto no es nada similar (o al menos no aún) a una inteligencia creativa o siquiera una mente y que más bien es un aparato regulador y automático como se ve en la siguiente cita:

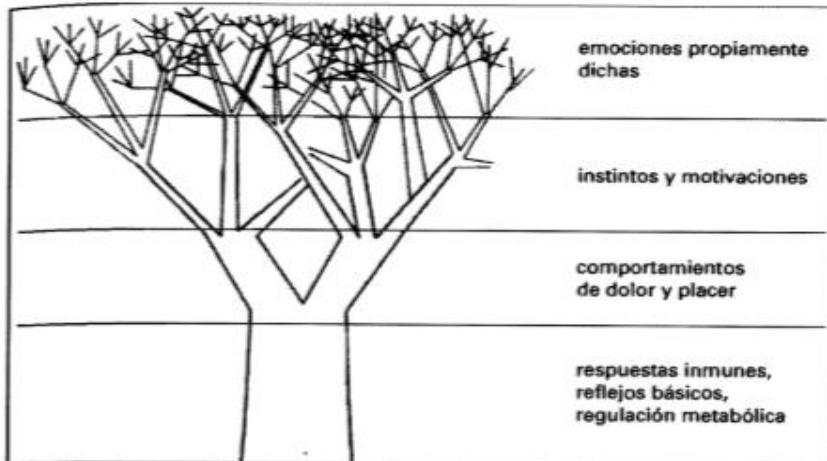
Todos los organismos vivos, desde la humilde ameba hasta el ser humano, nacen con dispositivos diseñados para resolver automáticamente, sin que se requiera el razonamiento adecuado, los problemas básicos de la vida. Dichos problemas son: encontrar fuentes de energía; mantener un equilibrio químico del interior compatible con el proceso vital; conservar la estructura del organismo mediante la reparación del desgaste natural; y detener los agentes externos de enfermedad y daño físico. La palabra

²⁷ Con emoción bidimensional doy a entender una emoción que funciona como estado fisiológico y mental a la vez.

homeostasis es el término apropiado para el conjunto de regulaciones y el estado resultante de vida regulada. (Damasio, En Busca de Spinoza, 2005)

Este mecanismo por lo menos en los seres vivos de hoy en día es innato e increíblemente refinado; mientras más refinado, mayor número de habilidades para realizar múltiples funciones tiene esta capacidad que en concordancia con Damasio llamaré “Máquina Homeostática”, en la base de esta máquina se encuentran las respuestas más básicas como: regular automáticamente la química interna; acercarse y alejarse de algún objeto; o bien aumentar o disminuir la actividad del organismo. La segunda etapa consiste en procesar las “sensaciones” de agrado o placer (recompensa) y desagrado, o dolor (castigo), esto se presenta como respuesta frente a un cierto cambio en el cuerpo que puede ser perjudicial o benéfico para el organismo, por ejemplo, un ser humano se sumerge en agua fría cuando hay altas temperaturas y siente placer, ya que puede disminuir el calor que sufre el cuerpo, en cambio si este sufre una quemadura, experimenta dolor, para alejarse de aquello que le resultó dañino. El nivel inmediatamente siguiente son los instintos y motivaciones, estos ya reflejan una capacidad de proyección de parte del individuo, y son por ejemplo el hambre, la sed, la exploración, el sueño, el juego (más propiamente el deseo de jugar para evitar el aburrimiento) o el sexo. Representan un grado de proyección, ya que el deseo se presenta frente a objetos no presentes en el entorno inmediato del animal. Previamente los he tratado como emociones, si bien no lo son propiamente, estas cumplen con las características que tienen las emociones: provocan pensamientos acordes con la sensación, conllevan cambios fisiológicos y son reconocibles mediante la experiencia, por lo cual para este texto los instintos serán tratados como emociones, pero sin ser considerados como emociones bidimensionales. El último grado son las emociones propiamente dichas, y que a continuación clasificaré entre bidimensionales y “unidimensionales”, la siguiente figura grafica el orden:

Fig. 5



(Damasio, En Busca

de Spinoza, 2005)

El neurobiólogo considera que las emociones propiamente dichas se clasifican entre tres posibilidades las cuales son: A) emociones de fondo, B) emociones primarias y C) emociones sociales.

A) Emociones de fondo: Son el resultado de los procesos homeostáticos de carácter inferior, la combinación específica de estos procesos lo llamaré como estado de máquina homeostática; estos estados se dan por periodos cortos, y suceden uno al otro, de tal modo que los seres humanos desde el momento en que nacen se encuentran siempre en un estado de máquina homeostática; los estados de máquina homeostática tienen como función llevar al organismo a corregir lo que esté funcionando mal en él. Además, estas emociones son aquellas que trabajan al momento en que a una persona se le pregunta “¿Cómo estás?” y ella responde “he estado bien, muchas gracias”, ese “bien” en realidad representa: no he sufrido dolores, no tengo grandes deseos sin satisfacer, tengo energía suficiente, entre otras. A este análisis que representa la respuesta “bien” se le llama “estado de ánimo” que se manifiesta de manera resumida en una sola emoción. En otras palabras, el “estado de ánimo” es la toma de conciencia de la emoción de fondo provocada por el estado de máquina homeostática. Las emociones de fondo tienen un carácter meramente fisiológico, es una manera de sentir (más cercana a la sensación que al sentimiento) el estado del propio cuerpo en vez de procesar cada factor en específico, una manera útil de diferenciarlas de los instintos es que estos suelen afectar un solo sector del

cuerpo, en cambio la emoción de fondo es sentir aquellos instintos que tiene el cuerpo y las sensaciones básicas de dolor o placer que provocan.

- B) Emociones primarias (o básicas): Estas son las emociones más fáciles de reconocer, se caracterizan por una clara notoriedad en el cuerpo, las más comunes son la tristeza la alegría, la ira, el asco, el miedo, la sorpresa, entre otras, una manera fácil de identificarlas es hacer el ejercicio de considerar todas las que se vienen a la mente cuando se escucha la palabra emoción. Esta notoriedad y consideración se debe a lo fácil que es diferenciarlo en los seres humanos y en otros animales, lo más común es que cuando una persona se encuentra en un estado de tristeza, su mirada se dirige al piso, su voz tenga un volumen bajo, entre otras características fácilmente reconocibles, y por lo tanto no se requiere tener una gran experticia para reconocer que otro experimenta una de estas emociones específicas. Como ya he dicho con el caso de la tristeza, estas emociones se pueden catalogar como bidimensionales, ya que tienen un carácter fuertemente mental.
- C) Emociones sociales: Estas incluyen la simpatía, la turbación, la vergüenza, la culpabilidad, el orgullo, los celos, la envidia, la gratitud, la admiración, la indignación y el desdén. Como subcomponentes de las mismas se podría identificar toda una comitiva de reacciones reguladoras, junto con elementos presentes en las emociones primarias, en combinaciones diversas. Piénsese de qué manera la emoción social “desdén” toma prestadas las expresiones faciales de “repugnancia”, una emoción primaria que evolucionó en asociación con el rechazo automático y beneficioso de alimentos potencialmente tóxicos. Incluso las palabras que se utilizan para describir situaciones de desdén y de afrenta moral (una expresión común es reconocer que el acto provoca asco o repugnancia) giran alrededor de este “anidamiento”²⁸. Ingredientes de dolor y placer son asimismo evidentes bajo la superficie de las emociones sociales, aunque de manera más sutil que en las emociones primarias. Las emociones sociales son fundamentales a la hora de llevar a cabo nuestras relaciones sociales, continuando con el desdén, esta determina las relaciones humanas a un nivel práctico, por ejemplo, entre dos compañeros de trabajo, donde uno le provoca desdén

²⁸ El anidamiento es el proceso mediante el cual un proceso simple genera, múltiples procesos más complejos, siguiendo la figura 5 el proceso de anidamiento corresponde al momento en que las ramas se dividen.

al otro, quizá logren llevar a cabo su labor de manera eficiente, pero difícilmente habría una relación cercana entre ellos, de tal modo es poco pensable que se compartan secretos o se inviten mutuamente a cenar un día a casa del otro; también tiene expresiones a niveles sociales más grandes, por ejemplo, los procesos democráticos de elección de candidatos. Estas emociones de nueva cuenta tienen un fuerte componente mental, si bien se dan debido a un estímulo más específico que las demás (difícilmente se dan por relaciones puramente materiales), la tristeza puede ser desencadenada por múltiples factores, desde el deterioro de algún objeto, un quiebre amoroso, mudarse de un lugar a otro, un objeto o lugar particular que tiene una carga afectiva o no poder satisfacer necesidades propias o de alguien con quien exista un lazo afectivo, en cambio las emociones sociales suelen dirigirse frente a actos humanos donde existe cierta voluntariedad de parte del EEC.

En resumidas cuentas, serán tratadas como emociones bidimensionales las emociones primarias y las emociones sociales, ya que no requieren que el EEC sea necesariamente material y por tanto aun cuando se presenten por una instanciación “material” específica, esta reduzca a la emoción a su característica materialista.

Como ya se ha dicho antes, Damasio considera que las emociones tienen como propósito regular la química interna del cuerpo, dando como resultado una mejor oportunidad de sobrevivir, y en vista de que existen emociones sociales, se puede interpretar que el cuerpo humano por lo menos reacciona químicamente ante las relaciones sociales que lleve la persona, además, como ya se ha visto en los ejemplos anteriores, las emociones primarias también pueden ser desencadenadas por nexos sociales, y según el principio de anidamiento estas desembocarían en las emociones sociales, por tanto, deduzco que el organismo²⁹ y sus nexos sociales tienen un cierto grado de dependencia. Por otro lado, el hecho de que las emociones se manifiesten con un output conductual, puede interpretarse como que aquellos que en su proceso evolutivo manifestaban la emoción que sentían de tal modo que otro ser humano pudiese reconocerlo, tenían una ventaja por sobre aquel organismo que se mantenía sin expresiones precisas. Llegado a este punto, es el primer

²⁹ En este punto en específico con “organismo” doy a entender un cuerpo sin mente como si estuviera hablando desde una postura cartesiana.

momento en que las emociones abren la puerta a un análisis sobre la adjudicación de estados mentales.

La capacidad de adjudicar estados mentales es aquello que permite que los seres humanos se desenvuelvan de la manera más óptima entre ellos, por ejemplo, si dos vehículos se cruzan por una calle que tiene una sola pista, de tal modo que uno de los dos debe subir a la berma para dejar pasar al otro, ninguno de los choferes podría tomar una decisión sin asumir que existe un conductor de la otra máquina, y que este a su vez es pensante, ya que si no fuese así, el vehículo contrario no se movería o por lo menos difícilmente lo haría de tal modo que fuera posible coordinarse. En este caso el *mindreading* de cada uno de los individuos deberá trabajar de la manera más óptima posible en intentar (ya habiendo hecho la adjudicación de estados mentales) predecir cuál será el comportamiento del otro chofer. Es probable que el comportamiento del otro genere alguna emoción, o incluso que la predicción que se realice sobre el comportamiento ajeno genere una emoción³⁰, pero esto no implica que exista un nexo crucial, es más, para el caso de las emociones sociales pareciese necesario haber hecho una adjudicación previa, de tal modo de reaccionar conforme o disconforme con los pensamientos o actos del otro, mientras que las emociones primarias no requieren de la presencia de otra mente.

Siguiendo esta línea, aún no se sabe nada sobre si el autómatas del experimento mental tiene o no estados mentales, solo es posible decir que tuvo un output exterior debido a algún proceso interno, pero no que esto se generase en conjunto con algún tipo de pensamiento o que el estímulo que provocó esa sonrisa fuese pensado, pero por lo menos en avance, es factible asumir que ante aquella sonrisa que lanza, su mirada de curiosidad o su baile, el espectador del material (o quien se imagina la situación) tiene una respuesta emocional propia y eso tiene algo que ver con la adjudicación de estados mentales. Caso distinto es el momento en el cuál la autómatas mira su mano “sorprendida” al notar que un trozo de madera la atraviesa por la palma; ella en ese instante abre los ojos de tal, manera que uno puede interpretar que analiza el daño sufrido, situación que es factible

Tra e³⁰ En la medida que las emociones son una función avanzada de la máquina homeostática y esta se dedica a regular la química interna del cuerpo y la reacción que este tenga con el medio, difícilmente a la edad necesaria para conducir habrá un objeto en el mundo que no genere una reacción emocional si es que este se hace consiente para el animal, por ejemplo, el mal aliento que pueda tener, no genera ninguna reacción emocional hasta el momento en que alguien se lo menciona o él percibe el olor que tiene.

considerar como un símil al proceso homeostático de los seres vivos, como ella carece de los procesos básicos (dolor y placer corporal), recurre a otro método para mantener su cuerpo en óptimas condiciones, dado esto último, sería posible adjudicar emocionalidad, puesto que la robotiza posee herramientas para tomar consciencia del estado de su cuerpo, sin embargo, en el mejor de los casos, esta sería una emoción de fondo y el proceso sería mucho menos eficaz que el biológico, sin despejar por completo la duda sobre su posesión de estados mentales. Para despejar la incógnita es necesario ahora revisar el nexo que existe entre una emoción y un sentimiento, ya que en esta relación se cimienta la relación del cuerpo y la mente y como las relaciones sociales que tienen un carácter más cercano a la mente que al cuerpo conectan con la emocionalidad. Finalmente, una vez que se tenga este aparataje se puede iniciar el contraste con las teorías clásicas (TT y TS). Lamentablemente este análisis solo clarificaría el funcionamiento en animales con un sistema neuronal similar al humano, dejando de lado el caso de estudio particular (el autómeta) ya que en caso de que las emociones cumplan un rol fundamental, después de toda esta explicación sólo se probaría que el criterio que utilizan los seres humanos para adjudicar estados mentales es emocional, pero esto no implica que el criterio para la presencia de estados mentales también lo sea, y aún si se probara que si lo es, no se podría afirmar que el robot posea emociones y por tanto tenga estados mentales; de todas formas una vez que los datos se encuentren sobre la mesa, es permisible ensayar una respuesta para esta problemática, a la luz del aparente caso de proceso homeostático que exhibe la autómeta al ser dañada.

Los sentimientos y la conexión entre mente y cuerpo

Volviendo a la entrevista que le realizaron a Damasio, en la cita anteriormente nombrada este explica que una vez percibidos todos los cambios que se dieron en el cuerpo se tiene el sentimiento, pero esto no resulta esclarecedor más allá de diferenciar al sentimiento como mental y a la emoción como fisiológica, la cita ya dada de la entrevista continúa:

(...) Así que percibiremos simultáneamente que alguien ha gritado (y eso nos inquieta), que nuestra frecuencia cardiaca y nuestro cuerpo cambian, y que, cuando oímos el grito, pensamos que hay peligro, que podemos o bien quedarnos quietos y prestar mucha atención, o bien salir corriendo.

Y todo este conjunto -el estímulo que lo ha generado, la reacción en el cuerpo y las ideas que acompañan esa reacción- es lo que constituye el sentimiento. Sentir es percibir todo esto, y por eso vuelve a situarse en la fase mental. De modo que empieza en el exterior, nos modifica porque así lo determina el cerebro, altera el organismo y entonces lo percibimos.
(Damasio, El cerebro, teatro de las emociones, 2008)

En este punto es necesario volver un momento sobre las emociones; como la cita ha dejado entrever, las emociones tienen un rol causal sobre nuestra mente, los pensamientos o imágenes que se tienen al momento de experimentar una emoción serán coherentes con esta misma, por ejemplo, si alguien se encuentra alegre, lo más probable es que tenga pensamientos (entiéndanse recuerdos o imágenes mentales) que estén en la misma sincronía que su alegría, tales pensamientos pueden ser algún momento en que experimentó una profunda alegría o la imagen mental de un lugar reconfortante o un deseo cumplido. Aquí se da uno de los fenómenos más curiosos de la emocionalidad, además de tener este rol causal frente a la mente, la mente también tiene un rol causal frente al pensamiento, piense el lector en alguna ocasión en que haya sacado una buena calificación después de una evaluación compleja, o cuando ha recordado algún familiar cercano que haya fallecido, lo más probable es que experimente una sensación de alegría o tristeza coherente con el recuerdo, siendo ahora la mente la causante de la emoción. Esta correlación en la cual la emoción y el pensamiento coinciden, es condición necesaria para el sentimiento, pues como Damasio explica, el sentimiento es la toma de conciencia del estado de máquina homeostática. Es importante tener esta aclaración en mente, pues de otro modo se corre el riesgo de confundir el sentimiento propiamente tal con la emoción o con las imágenes mentales; por una parte, las imágenes o pensamientos, no son equivalentes a un sentimiento, si lo fueran deberían presentarse siempre juntos y con relativa semejanza entre una persona y otra, si la característica por ejemplo del pensamiento que conlleva enterarse de la paternidad fuera el ser feliz, sería extraño que alguien se entristeciese o experimentara temor ante tal noticia, sin embargo estas son reacciones bastante frecuentes en muchas personas al enterarse de su próxima paternidad, y aunque se considere que en realidad los pensamientos son felices en la medida del individuo y el contexto, una persona no podría experimentar sobre el mismo pensamiento dos emociones distintas, pero es muy común encontrar relatos de madres o padres que al enterarse de que están en cinta se entristezcan o asusten y años después al recordar el momento sientan felicidad. Por otro lado, la confusión entre emoción y

sentimiento es más compleja de despejar, puesto que para la presencia de un sentimiento es necesario que existan ciertos tipos de emociones específicas. Resulta que una persona puede experimentar cierta emoción como “alegría” y a la vez sentirse sumamente mal, tal caso se podría presentar para una persona que sufre una enfermedad que le provoca un inmenso dolor todo el tiempo; un día este se encuentra con algún amigo que no ve hace bastante tiempo, en ese momento se emociona y experimenta alegría, luego de conversar con su amigo parten juntos a algún bar cercano y en ese lugar se quedan varias horas, de todos modos el dolor producto de su enfermedad no se detiene, y a la mitad de la velada el amigo le pregunta “¿Cómo te sientes?” a lo que él contesta: la verdad es que mal. No es que este sujeto teórico no esté alegre, muy por el contrario, la velada ha sido amena y ha tenido pensamientos alegres recordando anécdotas con su compañero de mesa, de todos modos, el estar bajo el efecto del dolor le ha impedido sentirse del todo bien, diferenciando la emoción del sentimiento, además ese sentir mal le debiese resultar inubicable en ningún lugar del cuerpo, es cierto, el siente dolor y probablemente este sí podría ubicarse en un punto en específico, pero el dolor es la causa de ese mal sentir, no el mismo, del mismo modo que el fuego es causa del calor pero no el mismo calor, *El sentimiento, en el sentido puro y estricto de la palabra, era “la idea de que el cuerpo se encontraba de una determinada manera”*. (Damasio, En Busca de Spinoza, 2005) Sin embargo, esta definición no se ajusta del todo al lenguaje y a la propia argumentación que he llevado hasta este punto, por lo cual es preferible considerar que una definición más apropiada es “yo me encuentro de una determinada manera”, el cambio de cuerpo a “yo” tiene tres motivos, el primero consta de que durante toda la argumentación he utilizado un lenguaje mentalista, donde he aceptado al menos en mi vocabulario la mente como diferente al cuerpo, (sin embargo no he pretendido en ningún momento defender el dualismo de substancias) y en segundo lugar, el sentimiento si considera en su “toma” de conciencia los pensamientos que tenga el individuo, conformándolo (al sentimiento) tanto el aspecto mental distinto a sí mismo y el estado corporal, finalmente el pronombre “yo” es el indicado, debido a que aún, según los datos que he presentado, no existe ningún motivo como para asumir que alguien más que el propio individuo tiene alguna manera de sentir lo que esté siente, muy por el contrario, si se ocupasen los pronombre “él” o “tu” implicaría que alguien más tiene acceso a esa toma de conciencia privilegiada para quien lo siente hasta ahora. La función del sentimiento es dar el siguiente paso homeostático para el individuo, o

en otras palabras, regular los procesos que cumple el organismo ya no frente a una necesidad determinada como el caso del instinto, o a una reacción compleja que tiene el individuo frente a un objeto o situación como el caso de la emoción, sino que el sentimiento regula el cuerpo casi en todos sus aspectos como un operador de trenes a cargo de una red completa.

La empatía

Retomando el punto anterior, los sentimientos son una toma de conciencia realizada por el individuo sobre el mismo, por lo cual el sentimiento no es un pensamiento común y corriente, si lo fuera este no tendría ninguna diferencia radical frente a una emoción u otra representación que se atravesase por la memoria o la imaginación del sujeto, por el contrario, el sentimiento es aquello que toma conciencia de estos y se “muestra” como un “resumen” de cómo está funcionando el individuo en su totalidad, una manera de resumirlo sería “yo siento mi auto-comprensión” por lo tanto el sentimiento puede ser considerado como una metarrepresentación de uno mismo. No obstante, al hablar de empatía se suele decir (por lo menos en el lenguaje ordinario) que un individuo siente lo que siente el otro, pero ¿cómo sería aquello posible? Para sentir a otro, el sujeto tendría que tener conciencia plena de los pensamientos, emociones, instintos y las demás respuestas homeostáticas que experimenta el otro, y aunque el tuviese las mismas respuestas, el sujeto no podría estar al tanto de que es otra persona, debido a que, si fuera consciente de ello, tendría pensamientos acerca de su condición como otra persona que la primera persona no tiene y por tanto su toma de conciencia sería diferente, de tal manera en que sólo si este fuera un “espejo” capaz de imitar a la perfección a su contraparte, olvidando su propia identidad, podría sentir lo mismo que la otra persona. Sin embargo, es cierto que los seres humanos tienen la capacidad de por lo menos coordinar sus emociones, de tal modo que si existe algún grado de afecto (entiéndase por afecto una conexión emocional recíproca entre dos personas) uno debiese experimentar emociones cuando menos similares al otro, en este sentido la empatía resulta fundamental para el ser humano, ya que sin un mecanismo que le permita relacionarse de la mejor manera posible con otros de su especie en el ámbito de sus estados anímicos, el organismo tendría dificultades para su supervivencia. Esto en realidad se debe a que el otro individuo cumple el rol de EEC frente a quien se emociona, de esta manera no significa que la persona que siente empatía esté sintiendo la tristeza de otra persona, sino que está sintiendo su tristeza causada por la de esa otra persona, por tanto, la empatía desde la postura de Damasio, no

sería sentir lo que el otro siente, sino que en realidad sentir acorde con lo que el otro siente, sin embargo, esta concepción resulta compleja de entender en su totalidad, cuestiones como por ejemplo ¿Cómo la mente sabe que siente el otro?, ¿Para qué sentir como siente el otro?, o ¿Qué relación existen entre la empatía y el minreading?³¹ Por lo cual a continuación profundizaré sobre este punto.

Para empezar esta explicación recurriré a la forma en que Damasio explica los sentimientos, él lo grafica con la metáfora del “mapa corporal” (Damasio, En Busca de Spinoza, 2005), volviendo al ejemplo del operador de trenes, el cuerpo y la mente serían el mapa de la red (en analogía al mapa corporal), y el sentimiento sería las decisiones que tome el operador en pro de mantener la estabilidad y el buen funcionamiento de la misma, tomando en consideración cada una de las variables en el mapa, pero decidiendo en base a la generalidad de este. Es en esta configuración del “mapa corporal” que Damasio propone las “simulaciones corporales” en las cuales la mente puede tomar ciertos datos y simular cómo se darían las condiciones si fuese ella quien se ve enfrentada a pasar por aquellas condiciones.

Citando a Damasio:

Piense el lector, por ejemplo, que le cuentan un horrible accidente en el que alguien resultó gravemente herido. Durante un momento puede sentir una punzada de dolor que imita en su mente el dolor de la persona en cuestión. El lector se siente como si fuera él la víctima, y el sentimiento será más o menos intenso en función de la dimensión del accidente o de la medida en que conozca a la persona implicada. El mecanismo que se presume que produce este tipo de sentimiento es una variedad que yo he llamado mecanismo del «bucle corporal como si». Implica una simulación cerebral interna que consiste en una rápida modificación de los mapas corporales actuales. (...) El cerebro crea de forma eventual un conjunto de mapas corporales que no corresponde exactamente a la realidad del momento del cuerpo. (Damasio, En Busca de Spinoza, 2005)

En otras palabras, si se sigue el ejemplo de la red, el operador funciona de tal manera que imagina una red distinta a la que tiene, anticipando cómo obrar en esos casos. Pero el caso humano va más allá, pues el ser humano funciona de tal manera que él realmente experimenta

³¹ Si bien se puede llegar a pensar que la primera pregunta, y la tercera son lo mismo, o por lo menos están tan estrechamente relacionadas que no vale la pena diferenciarlas, en realidad su gran distinción se funda en un mecanismo que se enfrenta a un aspecto meramente emocional, y la segunda atiende a la capacidad de “leer” no solo los aspectos emocionales, sino que los aspectos conductuales, las intenciones y decisiones de los demás.

las sensaciones que tendría en el caso particular, lo cual sumado a la palabra que en términos precisos es la correcta (simulación) que utiliza Damasio puede confundir e inclinar al lector a creer que se habla desde una postura simulacionista clásica. Para entender las diferencias, repasaré el factor emocional de la *teoría de la simulación*, está siempre tuvo un enfoque “emocionalista” en la medida en que consideran que los sujetos desde la muy temprana infancia suelen imitar a sus cuidadores, ya sea a un nivel emocional o por lo menos conductual, y en base a la experiencia acumulada de estas emociones, luego realizarían inducciones sobre los estados mentales de aquello que están imitando, para culminar con la aparición del mindreading entre los cuatro a cinco años de edad. Esto se hace cargo del aspecto individual que tienen las emociones, pues al ser corporeizadas deben ser tan variadas como las variaciones en los cuerpos entre un individuo y otro. Pero esta postura tiene el defecto de ver las emociones como instrumentos que utiliza la propia mente y no como agentes participes de esta, cuando el sentimiento de empatía se presenta con una contradicción, o en otras palabras, la persona que siente la empatía logra entender las emociones del otro, pero aun así entra en un estado de sentimiento diferente a la de ese otro, por ejemplo, cuando una persona se entera de que alguien que le provoca desdén sufre algún problema cuya intensidad coincide con la de su desdén, este podría sentirse feliz, la persona que experimenta la felicidad ha logrado comprender qué debe estar sintiendo el otro, su mapa interno tomó las características que tendría el mapa interno del otro y él ha experimentado sus emociones, sin embargo, las emociones del “simulador” no coinciden con las del “simulado”, y este experimenta a su vez otras emociones que difícilmente la persona que se enfrenta a la dificultad tuvo presente; si bien esto en una primera aproximación, no afectó su comprensión, los estados mentales que experimentará consecuentemente en la etapa de sentimiento impedirán su juicio, para un simulacionista clásico, por mucho que el individuo experimente emociones, o “imite” las que él cree que experimenta el otro (a nivel consciente o subconsciente) difícilmente esto debería afectar un resultado final del proceso, en el caso de Damasio por el contrario, no sólo es posible contar con explicaciones emocionales *of-line* (que el objeto no está presente) coherente con el pensamiento simulacionista, sino que la configuración es mixta, en otras palabras el sujeto realiza la “inducción” (si es que hay una)

teniendo en cuenta tanto EEC *of-line* como EEC *on-line* (las emociones que siente hacia su propia simulación)³².

Volviendo a la explicación de la empatía, los “mapas corporales” son el mecanismo que utiliza la mente para ponerse en concordancia con las emociones del otro, pero estas solo consisten en la mitad de la respuesta a la primera pregunta, aún Damasio se enfrenta al problema *mente-cuerpo*, es aquí donde él propone como posible respuesta a las “neuronas espejo”, en el texto *Evidencia de actividad de la neurona espejo durante la observación de acciones humanas y de robots: hacia un Análisis de las cualidades humanas de los robots interactivos*³³ (Oberman, McCleery, Ramachandran, & Pineda, 2007) se explica que las “neuronas espejo” son un conjunto especial de neuronas que poseen algunos primates, entre ellos los seres humanos, las cuales al captar algún objeto que el cerebro reconoce como similar a sí mismo, se activan con tal de propiciar una imitación conductual y en el caso de los humanos emocional, del mismo modo funcionan tanto de manera *on-line* como *off-line*, siendo capaces (los primates) de imitar lo que otro hizo mediante estas neuronas tanto en el momento que lo observa, como en el momento que lo recuerda. No por esto se debe interpretar que las neuronas espejo imitan a la perfección a otros seres, más allá de las capacidades fácticas que tenga un primate con respecto a otro para el movimiento. Ya he fijado que resulta imposible experimentar las emociones ajenas, y que estas son propias, el aporte de las neuronas espejo en cambio, es explicar que existe un mecanismo el cual se dedica sólo a la revisión de los otros y decide por sí mismo qué emoción adjudicarle a los otros, con tal de manifestar la propia versión en uno mismo generando el “mapa corporal”.

El segundo problema a tratar es la utilidad de la empatía, como ya he esbozado durante el texto, resulta útil para las relaciones sociales, el comprender el estado emocional de los otros, pero ¿por qué? La respuesta recae nuevamente en la evolución, a diferencia de otras especies, los seres humanos tienen grandes debilidades en su relación con el medio, por nombrar algunas de las más evidentes se encuentran: su gran tamaño, lo cual requiere grandes cantidades de energía; una infancia muy prolongada, siendo capaces de sobrevivir relativamente bien en el medio recién entrando a la adolescencia a eso de los catorce años;

³² Próximamente dentro de este mismo capítulo volveré a abordar una refutación al simulacionismo clásico desde una postura de Damasio.

³³ Traducción propia.

baja velocidad de carrera y poco tiempo manteniéndola; cabezas muy grandes en la infancia, y ausencia de “armas” incorporadas frente a otros animales. Todas estas dificultades fueron paleadas en un principio mediante la congregación social, por ejemplo, a la hembra de la especie le convenía quedarse con el macho progenitor, ya que mientras ella se dedicaba al cuidado de la cría, él podía reunir alimento o protegerla de depredadores. La empatía particularmente resulta útil en esta dinámica prehistórica, entender de algún modo el sentir emocional del otro, aseguraba que se tomaran medidas en conjunto para solucionar el problema que se presentara y mantener la unidad del conjunto humano. Por ejemplo, si alguno de los padres logra experimentar frente a un estímulo específico el malestar de su cría, este puede correr para socórrerlo e intentar solucionar su conflicto, es aquí donde se da el nexo entre el mindreading y la empatía. De cierto modo la empatía da una comprensión relativamente cercana sobre los sentimientos del otro, mientras que el mindreading da una comprensión sobre los estados mentales, y ya que el sentimiento es un estado mental, la empatía debiese ser una habilidad propia del mindreading, sin embargo, existe un estudio que refuta esta inferencia, el texto *Relaciones entre empatía y teoría de la mente en niños y adolescentes* (Zabala, Richard's, Breccia, & López, 2018) somete a estudio a un grupo de niños y adolescentes; aunque toma como presupuesto que la empatía y la ToM³⁴ tienen que ver con los otros, son esencialmente distintos ya que la ToM no se encarga de la emoción más allá de un entendimiento representacional y la empatía es incapaz de generar comprensión frente a cualidades intencionales, representacionales, etcétera.

La empatía y la TM son procesos similares, por cuanto permiten a las personas establecer vínculos a partir de la comprensión de los estados mentales y emocionales de los otros. Sin embargo, el estudio de las bases neurales implicadas en ambos procesos ha generado suficiente evidencia neurocientífica que indica que estos poseen correlatos neurales diferentes. Estos estudios demostraron que ambos procesos implican la activación de sistemas neuronales comunes en lo que respecta al componente cognitivo de la empatía, y al mismo tiempo, la capacidad empática involucra correlatos neurales específicos asociados al procesamiento emocional. Así, la ToM afectiva asiste a la capacidad empática en el procesamiento de la información relativa a la atribución de emociones y a la comprensión del estado emocional experimentado por las otras personas. De este modo,

³⁴ Producto de la inclinación hacia la teoría de teorías del estudio, me referiré al mindreading como ToM dentro de este contexto

ambos procesos se interrelacionan, dando lugar a la capacidad de responder a los estados emocionales ajenos, tanto desde la comprensión como desde la posibilidad de experimentarlos. (Zabala, Richard's, Breccia, & López, 2018)

En el estudio que ellos produjeron para probar la hipótesis citada, encontraron que a medida que los sujetos de estudio se encontraban más cerca de la adultez, menos dependientes se mostraban de las zonas del cerebro implicadas en sus capacidades de ToM afectiva, y veían un aumento en sus capacidades empáticas. Mientras que los niños menores, recurrían con mayor frecuencia a un margen representacional para entender el estado mental emocional ajeno, por lo cual, según ellos, la ToM y la empatía funcionan en márgenes distintos. Esta postura es cierta en la medida que se considere a la mente como divisible en “partes”, coherentemente según sus funciones y zonas del cerebro que trabajan en cada una de ellas, pero como veré más adelante, esta aproximación es cuando menos errada, si bien es cierto que un daño en algún sector del cerebro puede generar un mal funcionamiento mental, o incluso la pérdida de una capacidad de esta misma, la mente sigue trabajando sin ella y rellenando el espacio perdido de la mejor manera posible; esto solo prueba que ella (la mente) funciona en completitud, teniendo en cuenta el cerebro, el cuerpo y la mente como uno. Volviendo al caso del operador de la red de trenes, la propuesta de estos sujetos solo implica que llegado determinado momento, el operador centra su atención en ciertas líneas o estaciones con la finalidad de hacer más eficiente la red, pero esta no podría trabajar de modo normal sin alguna de sus partes o el operador, en otras palabras, esta distinción sólo funciona en un margen de estudio, por tanto no podría trabajar en la realidad de manera óptima, pues del mismo modo que la red de trenes no es la suma de sus estaciones (sino un objeto único), la mente funciona normalmente en completitud. Esto me autoriza a un nivel filosófico a considerar esta opción como poco relevante para mi hipótesis particular.

Phineas Gage, las teorías clásicas y la intersubjetividad

Uno de los casos más interesantes para la ciencia cognitiva es el caso del enigmático Phineas Gage. Él fue un hombre modelo para su época, sin ninguna dificultad en sus relaciones sociales y muy educado en normas de conducta, responsable trabajador y eficiente, si es que alguien hubiese previsto su futuro probablemente no creería que moriría solo en la locura después de años de miseria, pero debido a su personalidad carismática, sociable y fácil liderazgo lo que resultaría más impredecible sería su soledad, acompañado únicamente de

una hermana que lo cuida por caridad más que cariño. Lo curioso de Gage no fue tanto su locura, pues esta no era como generalmente se piensa en un hombre demente, él no tenía ninguna de sus capacidades disminuidas, su memoria se mantuvo intacta hasta el final, sus habilidades frente al pensamiento lógico y las operaciones matemáticas tampoco fue afectada, sus razonamientos eran acertados, su lenguaje y capacidad de comunicar sus propias ideas o entender las ajenas fue normal hasta el día de su muerte, al menos en lo que respecta a las capacidades sintácticas y semánticas. La locura de Gage era predominantemente social, no porque no tuviese capacidades sociales, sino porque un fatídico día él perdió una emoción, convirtiéndolo en un ser sin vergüenza en toda la norma.

Gage trabajó como capataz de construcción para los ferrocarriles Rutland y Burlington, cargo que producto de su responsabilidad y compromiso con el trabajo ya ejercía a sus 25 años, y había demostrado una increíble capacidad de liderazgo que pronosticaba que iba en ascenso y pronto sería promovido, pero en el verano de 1848, Gage se encontraba realizando una tarea usual para él, detonar con dinamita piedras que se atraviesan por la trayectoria del ferrocarril en construcción a través de un sistema que requería el uso manual de una barra de metal, (herramienta especial que el mismo había mandado a construir), sin embargo, a media tarde un compañero lo distrae por un instante, el descuido produjo que la dinamita detonara antes de tiempo, disparando su instrumento (la barra de metal) a través de su cráneo, perforando el cerebro y reventando uno de los ojos del pobre hombre. Lo más curioso del accidente fue que a pesar del daño, Gage no perdió el conocimiento, se mantuvo lucido todo el tiempo, y a pesar de que la barra había perforado toda su cabeza, la recuperación fue bastante rápida (no más que un par de meses), pero el ya no era el mismo, porque a pesar de que su cuerpo sufrió varios daños, las secuelas físicas no se comparaban a las psicológicas. El médico que lo trató no notó estos cambios hasta adentrados ya muchos meses, en una ocasión en que Gage le cambió el nombre, sólo a modo de jocosidad igual que un niño; cuando le habían hecho pruebas y test psicológicos rudimentarios debido a la época, pero útiles por lo menos, los había pasado sin ningún problema, por lo cual no parecía que su intelecto se viera mermado por el daño cerebral, pero ahí un hombre adulto se estaba mofando de él al igual que un niño. Cuando Gage fue devuelto a su trabajo en su puesto, primero empezó a mofarse de sus compañeros y su dirección se volvió ineficiente e incómoda para sus colegas, su productividad personal, junto con la responsabilidad al enfrentar las cosas que

se le pedían era casi nula, sin embargo sus conocimientos y capacidades físicas para realizar la obra estaban intactos, por lo cual sus empleadores trataron de retenerlo, dándole trabajos cada vez más ligeros y menos importantes, pero su desinterés en la convivencia y el trabajo mismo hicieron imposible su permanencia. Con el paso del tiempo esta situación solo se agravó, Gage viajó por toda América buscando oportunidades laborales, pero su interés solo en sus deseos inmediatos le imposibilitó mantenerse en uno, literalmente el solo hacia aquello que le complaciera, además solía insultar al resto de las personas a su alrededor, a tal punto que a aquellos que lo conocían por primera vez se les advertía de su conducta. El daño específico en su cerebro le impedía procesar una emoción en particular, “la vergüenza”, si bien Phineas Gage era consciente de que sus actitudes estaban éticamente erradas, la respuesta negativa que genera la vergüenza al experimentarse había desaparecido, dejando a un ser totalmente distinto de lo que había sido, por lo demás, el procesaba sin ningún problema la mayoría de las emociones y las reconocía en los demás, ya que intentaba inducir aquellas que le resultasen agradables en aquel momento, por lo cual su capacidad de generar mapas corporales seguía ahí, pero su ToM no funcionaba de manera correcta, ya que el propósito social de esta no podía llevarse cabo. Para cualquiera de las teorías sobre ToM clásicas, explicar a Gage resulta imposible, el posee elementos suficientes como para encajar en el perfil que propone cada una de estas, y aun así él no funcionaba como debería, siendo un contraejemplo útil para ambas posturas, pero a esta altura es necesario sumar la postura intersubjetiva o de segunda persona, la cual es una alternativa a las dos anteriores.

1) Teóricos de teorías:

La postura más débil a la hora de enfrentarse al caso Gage es la teoría de teorías, el hecho de que defendiesen un modelo cercano al computacionalismo les resulta cuando menos poco favorable. Para Cohen, la emocionalidad y el reconocimiento de emociones es un módulo particular en la mente humana, donde las emociones y sentimientos de otros se trabajan como representaciones computables, que son proveídas por dos módulos menos complejos (Zilber, 2017), aunque luego se agregaron dos módulos más, el módulo de empatización (Empathizing) y el módulo de detección de emociones (Zilber, 2017). En la medida en que la emoción de un sujeto se muestra en el cuerpo, la mente humana no debiese tener problemas

en procesar las emociones ajenas en forma de representaciones y luego computarlas incluso para que el cuerpo del sujeto las experimente. El problema con esta teoría al enfrentarse a Gage es la falta de “sentido común” del sujeto; al perder parte del cerebro en el accidente y con ello la capacidad de sentir vergüenza, la habilidad de llevar a cabo de manera efectiva sus relaciones sociales se vio afectada. Si es que realmente la mente humana y particularmente la adjudicación de estados mentales fuera modular, persistiría el problema de ¿Cómo es que Phineas Gage puede hacer una parte de la adjudicación de estados mentales y ser incapaz de actuar acorde con la empatía que sentirá? ³⁵

Si se considera que la mente es serial, esta no podría dar un paso específico para que en la mente de Gage sea preferible lanzar un insulto y no callar su deseo, ya que ambas tareas deberían ser llevadas a cabo por el mismo módulo y en caso de este fallar, Gage no podría llevar a cabo uno de los pasos que le hacen tomar la decisión que tomó. Por otro lado, si la mente fuera de modulación distribuida si se podría aceptar que Gage funcionará con relativa normalidad como lo hizo, pero fallando en la decisión correcta, aunque la teoría fallaría, puesto que le sería imposible explicar de qué manera solo se pierde la emoción “vergüenza”, procesando aún todas sus otras emociones y procesos lógicos, en la medida que para ello necesariamente Gage sigue procesando metarrepresentaciones, pero aun así no puede representar la vergüenza.

Frente al caso del autómata, a decir verdad, la teoría no tendría muchos contratiempos a la hora de explicar la adjudicación de estados mentales, ya que el autómata cumple con la mayoría de las características observables a simple vista que tendría una mujer común y corriente, por tanto, considero innecesario profundizar en una explicación sobre este punto.

El segundo problema de la teoría de teorías es que intenta dar una explicación de las emociones y los sentimientos en la cual estos sólo son considerados como un aspecto más que debe ser contestado para realizar una adjudicación de estados mentales precisa, pero no considera que las emociones puedan tener algún rol en esta toma de decisiones. Además, en la medida en que las emociones son provocadas por un EEC y que todos los objetos con los

³⁵ Pareciera ser que Phineas Gage realizaba la adjudicación de estados mentales “a medias”. Por una parte, él sabía que el otro pensaba, pero era incapaz de sentir empatía en la medida que su mapa interno no se adaptaba del todo al mapa interno ajeno.

que el individuo interactúa en el mundo son EEC, a los seres humanos les sería imposible en algún momento estar “trabajando mentalmente” sin una emoción, y si la mente humana operara sin el rol causal de sus emociones, los módulos deberían estar trabajando a la par entre procesar la adjudicación de estados mentales y las propias emociones, por tanto el módulo que se dedique por ejemplo a la tristeza debería procesar y a su vez entender la tristeza ajena y la propia.

2) Teoría de la simulación:

Como ya he dicho en este capítulo, la teoría de la simulación siempre ha considerado la emocionalidad como un factor, a diferencia de la teoría de teorías los simulacionistas no ven a la emocionalidad como un factor más que debe ser explicado, en cambio, las emociones son parte de la simulación que realiza el individuo en tanto es propia de cada uno. Si se retoma la pregunta ¿Qué haría yo? Típica del simulacionismo, la versión que se haría para el caso de la emoción sería ¿Qué emoción tendría yo?

Pero el alcance que aún no llega a hacer la teoría de la simulación, radica en cómo operan las emociones a un nivel práctico más allá del teórico; las emociones en la teoría de la simulación son herramientas dadas a la mente tal como podrían ser sus capacidades lógicas o la memoria (doy a entender con esto la capacidad de recordar y no los recuerdos). En cambio, siendo coherente con el sistema que ya he propuesto a lo largo de este texto, la emocionalidad no solo es una herramienta para regular el cuerpo, si bien ese es su propósito particular, en este mismo proceso es un regulador de esta misma y además parte de ella. Con esto anterior quiero decir que la mente es regulada por la emoción en tanto las emociones causan estados mentales; es constituida en parte por la emoción, ya que existen emociones bidimensionales y es el mecanismo que le permite mantener vivo a su “recipiente” dándole una mejor oportunidad de sobrevivencia, mejorando la posibilidad de entendimiento con los otros, pero teniendo a la emocionalidad como necesaria en este proceso (ya sea porque influencia la toma de decisiones, o porque ella misma “decide”), a diferencia de la teoría de la simulación que pone el centro en la experiencia del individuo, dejando de lado la situación presente de quien sufre la emoción.

La teoría de la simulación tiene dos problemas a la hora de enfrentarse a Gage. En primer lugar, está la imposibilidad de simularlo, ya que las simulaciones dependen en gran medida

de la propia experiencia, y muy pocos seres humanos aparte de Gage han tenido una lesión similar, por lo tanto sería difícil que alguien pudiera recurrir a su experiencia para entender las actitudes de Gage. El segundo problema radica en la importancia que se le brinda a la experiencia en esta teoría, ya que Gage antes del accidente había experimentado la vergüenza, él la conocía a la perfección, pero después del accidente no era capaz de procesarla de ninguna manera. Si Phineas Gage realmente simulaba a los otros ¿Por qué no a sí mismo antes del accidente? Gage tenía todas sus capacidades lógicas intactas, sus recuerdos y si se lo proponía era capaz de trabajar tan hábilmente como antes del accidente, por tanto, se puede pensar que si él se lo hubiese propuesto, habría engañado a todo el mundo y continuado con su vida, si no lo hizo es porque ya no podía simular en el sentido de Damasio, pero sí tenía las herramientas en él de los simulacionistas. Por una parte, la argumentación que he propuesto en base a Damasio implica que la simulación es al tiempo que la emoción, lo que significa que se presenta fisiológicamente, por tanto, Gage por mucho que recordara que la alguna vez sintió vergüenza, es incapaz de recordar como esta se siente, ya que es incapaz de entrar en el estado de máquina homeostática específico de la vergüenza. En cambio, para la teoría simulacionista clásica que acepta estímulos off-line sin implicación de una cualidad on-line, Gage no debería tener problemas para simular en su mente y actuar acorde a ello en la medida que su volición lo permita.

Frente al caso del autómatas, la teoría de la simulación tiene menos oportunidades de presentar una explicación; nuevamente la oportunidad de simular ¿Qué es ser como un autómatas?³⁶ Se ve impedida al enfrentarse a una máquina que carece por completo de un sentido como el tacto, no tiene recuerdos de la infancia, no siente hambre o sed, entre otras cualidades que sí tienen los seres vivos. Nunca ha habido alguien que cumpla todas estas características, y si es que lo hubo, yo no he escuchado sobre él, nada en mi propia experiencia podría socorrerme en este caso, pero esto no implica que la teoría de la simulación quede descartada. El video que he presentado muestra un robot muy parecido a los seres humanos (de hecho, es una actriz) por lo cual un cerebro humano puede reaccionar frente a ella como con cualquier persona que se le cruce por el camino, así que realizará sus funciones de simulación usuales hasta que ocurra algo que el sujeto no haya experimentado nunca, donde el sujeto está

³⁶ Esta es una referencia al famoso experimento mental “¿Qué es ser como un murciélago?” de Thomas Nagel.

validado a decir “no sé lo que se siente tal...” sin pasar a llevar la plausibilidad de que los seres humanos sean “simuladores”. En otras palabras, la teoría simulacionista falla por el margen explicativo acotado que le provoca el no considerar la simulación como un EEC, en contraste con la teoría de teorías que por su presupuesto racionalista no puede aceptar un fallo en su teoría sin proponer una teoría ad-hoc.

3) Teoría de la intersubjetividad

La teoría de la intersubjetividad o como también es llamada “perspectiva de segunda persona” surge como una alternativa a la teoría de teorías, a la par de la teoría de la simulación. En palabras de Antoni Gomila un buen resumen sería:

Invito a considerar la existencia de la perspectiva de segunda persona de la atribución mental, como una perspectiva diferenciada de las de primera y tercera persona. Su ámbito específico sería el de las atribuciones espontáneas y recíprocas en situaciones de interacción cara a cara, por lo que supondría su naturaleza expresiva. Al tratarse de la perspectiva ontogenéticamente primaria, ofrece una vía para superar las dificultades complementarias de los enfoques teóricos (tercera persona) y empáticos (primera persona) dominantes. (Gomila, 2002)

Esta postura asume como casos paradigmáticos aquellos en los cuales los sujetos están cara a cara y pueden interactuar con relativa normalidad entre ellos. Esto se debe a que la teoría de la intersubjetividad sí toma en cuenta las emociones en el sentido en que las he tratado (con un rol causal permanente) y de este modo (cara a cara) el sujeto puede leer de manera más óptima la emociones del otro. Pero la consideración no llega sólo a este punto, en realidad la teoría de la intersubjetividad considera que los dos sujetos provocan emociones entre sí, interactuando de tal modo que uno es dependiente del otro (al menos en la composición de sus estados mentales): las expresiones faciales, el lenguaje corporal la posición del cuerpo, etcétera, son presididos de modo directo y significativo, no como síntomas de una emoción o estado mental, sino como partes constitutivas de la misma, así mismo no requiere de algún tipo de metarrepresentación.

En tanto, la interacción directa propuesta, resulta favorable en la medida que la teoría de la intersubjetividad tiene un carácter eminentemente práctico y su intención no es centrar la discusión en los marcos teóricos, sino, en los márgenes prácticos de la vida cotidiana donde los sujetos se desenvuelven con un carácter expresivo, que a diferencia de los constructos

teóricos donde se tratan principalmente entidades no observables, la expresión de emociones es fácilmente observable. Pero ¿por qué llamarlo intersubjetivo? *Tradicionalmente, al concepto de Intersubjetividad se lo suele utilizar para dar cuenta del hecho de que una subjetividad toma a otra como foco de atención* (Martínez, 2011) por tanto, la relación entre los dos individuos interactuantes, en la medida que uno de ellos centra su atención en el otro y viceversa, logran coordinar (dentro de lo posible) sus estados mentales, siendo esto aquello que da nombre a esta corriente.

Sin embargo, esta teoría da un paso más allá de lo que me es legítimo aceptar, la teoría intersubjetiva al poner el ahínco en las relaciones emocionales que se provocan entre los sujetos, no logra hacerse cargo de la complejidad de estas en el fuero interno. Por mucho que las emociones se manifiesten de manera visible y que los nexos sean inmediatos, el ser humano (no se sabe si es que otros seres también) tiene un mundo interno no necesariamente manifiesto en la corporalidad. Con esto quiero decir que si bien existe un nexo causal entre las emociones que sienten mutuamente, no implica que exista una conexión entre ambos sujetos (ambos sienten sus propias emociones y no las del otro) imposibilitando un contacto real entre ellos. El teórico de la intersubjetividad se defendería diciendo algo así como: “Cuando un individuo ve a otro llorar, no piensa: Oh!, está botando lágrimas, gimoteando y tiene sus ojos irritados, por tanto debe estar triste. Muy por el contrario, pareciera ser que sabe directamente que el otro se encuentra triste, saltándose el razonamiento sobre los estados mentales ajenos”, pero una defensa de este tipo sólo reafirma mi posicionamiento desde la perspectiva de primera persona (no simulacionista) debido a la interpretación de los datos que ofrece el intelecto del espectador, quien finalmente entenderá los datos recopilados acorde consigo mismo y no con el otro; la información dada, propicia una inducción³⁷ de parte de quien observa que puede confundir la causa correcta, en este ejemplo, podría no estar llorando por tristeza, si no que por dolor o alegría, y aunque el simulacionista tiene como caso paradigmático la interacción directa entre

Los sujetos, no se puede obviar que la adjudicación puede ser llevada a cabo aún sin interacción presencial, como en el caso de la mensajería instantánea vía chat.

³⁷ Hablo en términos de "inducción" por fines explicativos, pero me inclino más por una concepción de conocimiento directo.

El caso específico de Phineas Gage resulta de poca utilidad, ya que este no era capaz de procesar una emoción en particular, cosa necesaria para poder ser analizado desde la postura de los teóricos de la intersubjetividad al ellos aceptar que los individuos en su acto comunicativo requieren estimular la emocionalidad ajena, por tanto, alguien que defienda esta perspectiva podría asumir que Phineas Gage o no posee mindreading o que este funciona mal.

Por otro lado, el autómata es un caso distinto; en vista del tipo de material que he trabajado, propongo la siguiente cita de Gomila para explayarme:

Para obtener una cierta base intuitiva de lo que entiendo por perspectiva de segunda persona en el ámbito de la atribución de emociones, quisiera apelar a su experiencia de espectadores cinematográficos, puesto que el cine trabaja con frecuencia presuponiendo múltiples mecanismos psicológicos, no sólo perceptivos. En particular, quisiera centrarme en la dimensión de implicación emocional que muchas películas explotan con tanta eficacia, justamente para tratar de aclarar lo que está implicado en tales procesos. Utilizando una terminología técnica procedente de la pragmática del lenguaje, podría decirse que algunas escenas fílmicas constituyen también actos perlocutivos, al buscar intencionadamente una determinada respuesta emocional por parte del espectador. A veces, esa respuesta de la audiencia depende de ciertos efectos «especiales» (sonoros, lumínicos...), o de la situación epistémica de los personajes (víctimas de engaños, confusión o desconocimiento de lo que les aguarda); paradigmáticamente, sin embargo, el efecto se produce en base a la visión de una escena de interacción humana (o sobre modelos humanos), cuyos actores casi inevitablemente experimentan emociones (o simulan experimentarlas, según la escuela de actores a que pertenezcan). Lo importante de este tipo de casos es que esta reacción emocional de los espectadores depende de su atribución de ciertas emociones al personaje, atribución que tiene lugar de un modo implícito. (Gomila, 2002)

En contraste, la actriz intérprete intenta todo lo contrario, dar a entender que no posee emociones, o al menos no emociones visibles, esto tiene un propósito de parte del director, provocar la sensación de que algo funciona mal para que los instantes que si den a entender emociones conlleven un mayor impacto. Pero dado esto último, por mucho que los espectadores le adjudiquen estados mentales al autómata, no han logrado probarlos, en otras palabras, la teoría intersubjetiva sólo es capaz de proponer un método, pero no un criterio claro, para efectos de este texto trataré como método a las “neuronas espejo”; *Estas neuronas,*

a las que se ha denominado «neuronas espejo» (*mirror neurons*), forman parte de un sistema percepción/ejecución de modo que la simple observación en el otro de movimientos de la mano, de la boca o del pie activa las mismas regiones específicas de la corteza motora como si el individuo estuviera realizando esos movimientos, aun cuando esta activación motora no se transforme en un movimiento actuado visible en el individuo que observa. (Tellez-Vargas, 2006). Pero cabe la pregunta ¿Cómo es que las neuronas espejo se activan con algo que se sabe es una máquina? En el artículo *Evidencia de actividad de la neurona espejo durante la observación de acciones humanas y de robots: hacia un Análisis de las cualidades humanas de los robots interactivos* (Oberman, McCleery, Ramachandran, & Pineda, 2007) se presenta como opción la posibilidad de crear a los robots con similitudes suficientes a la a un error argumentativo, sino, que, a factores para engañarlo, o en otras palabras, hacer al robot con un comportamiento y apariencia tan similar al humano que su cerebro sea inevitablemente engañado.

Llegado este momento repasaré los motivos por los cuales no adhiero a la teoría de la teoría intersubjetiva:

- 1) Le da una excesiva importancia al “otro” considerándolo como el foco de atención para el mindreading.
- 2) Desconoce como factor relevante las cualidades internas de la emocionalidad que conlleva en la mente humana.
- 3) Nada implica que exista un nexo necesario entre conducta observable y emoción (en la medida de que, por ejemplo, los actores logran regular sus respuestas faciales).

A modo de conclusión de este capítulo, dejo en claro que no he contestado el nexo que existe entre el Mind Reading y la empatía con los ejemplos dados; esto último no responde a la casualidad o a un error argumentativo, sino, que a factores más bien, metodológicos que abordaré en el capítulo 5.

Capítulo 5: una aproximación emocional al método y el criterio de adjudicación de estados mentales

El criterio emocional

como ya he dicho en reiteradas ocasiones: “el primer criterio para la adjudicación de estados mentales debe ser la emocionalidad” o, en otras palabras: “si es que otro tiene emociones por tanto debe tener estados mentales”; Ya se ha dicho que existen emociones bidimensionales, y que estas además de ser estados fisiológicos son a su vez estados mentales, cosa que podría ser considerada como suficiente para crear la emocionalidad como un criterio para esta adjudicación, sin embargo, mi apuesta da un paso más allá de tal modo que sitúa a las emociones a la par de otros criterios tales como el lenguaje y el comportamiento inteligente, en cambio mi propuesta dice que el factor emocional debe ser predominante, de tal modo que si algún objeto posee lenguaje y comportamiento inteligente, pero en cambio no presenta ningún tipo de emoción no puede ser llamado con certeza como un ser dotado de estados mentales, no así, si es que el ser del que se habla está dotado de emociones, pero pareciera no llevar un comportamiento inteligente o poseer un lenguaje, en este caso me creo completamente validado (según mi primera hipótesis) a aceptar que aquel ser si tiene estados mentales.

Tal aproximación parece cuando menos poco intuitiva, la tradición ha tenido buenos argumentos a su favor, y el factor fisiológico de las emociones propicia una toma de distancia frente al funcionamiento de los sistemas más apegados a la tradición cartesiana, los cuales suelen concebir a la mente como separada del cuerpo (por lo menos a nivel metafísico o explicativo). Además de esto no se puede olvidar que el hecho de que las emociones sean causales de otros estados mentales no implica necesidad, en otras palabras, el hecho de que exista emocionalidad a-priori no implica que existan estados mentales en aquel objeto, por tanto, la única manera de probar mi punto sería probar que existe necesidad entre la emoción y algún estado mental, de tal modo que siempre que se presente una emoción se presente consigo algún tipo de estado mental diferente a esta.³⁸

³⁸ Solo por fines retóricos, no tratare a las emociones como estados mentales, aun las que he considerado como bidimensionales.

Encuentro en los recuerdos el estado mental que necesito, pues la mera identificación de una emoción requiere de este sistema. Si se supone una emoción específica, ya sea tristeza o alegría, para su oportuna identificación como una u otra el ser vivo en específico debe por lo menos recordar la sensación adecuada coherente con la emoción que siente, pero la memoria no se detiene ahí precisamente, en cambio, ella es bastante influyente, pues es la encargada de brindar al pensamiento su material; *el pensamiento está constituido principalmente de imágenes* (Damasio, El error de Descartes, 1996). Pero estas imágenes no son replicas idénticas del objeto, por el contrario, el cerebro crea ciertas representaciones, lo más parecidas posible al objeto original. ahora bien ¿Cómo es que el cerebro jerarquiza que cosa debe ser privilegiada para almacenar de la mejor manera posible? El factor es emocional, mientras más intensa sea la emoción más clara será la representación. Esto se da a tal punto que incluso en casos amnesia o recuerdos bloqueados pueden recuperarse aquellos recuerdos con alguna emoción en particular. La necesidad en este caso no se ve de primer vistazo pues, aunque un recuerdo lleve consigo una emoción, una emoción no tiene por qué llevar un recuerdo. Aun así, el que la emoción sea el criterio que utiliza la mente para decidir o no resguardar una imagen, implica un proceso de asociación entre ellos, cosa que me faculta a pensar que una emoción conlleva un recuerdo en la medida que estas (las emociones) serán interpretadas en la medida que surjan recuerdos necesarios en ellas, por ejemplo, el caso de la alegría, esta no podría entenderse como alegría (y por tanto cumplir su función homeostática) sin proveer un recuerdo contrastable con la imagen mental presente. Esto se debe a que para formular un sentimiento debe ser catalogada la emoción, y para catalogar la emoción se requiere de un recuerdo, y en la medida de que las emociones necesariamente implican un sentimiento, me creo validado para aceptar este punto como suficiente. sin embargo, aún persiste el problema en objetos no humanos. Retomare el caso del autómatas para explicar cómo es que este sistema podría ser aplicable a inteligencias artificiales.

El caso de la máquina que aprende por Alan Turing es uno de los mejores, el compara a una IA:

En vez de tratar de producir un programa similar a la mente adulta, ¿por qué no tratar en vez de producir una que simule la mente de un niño? Si ésta fuera luego sujeta al curso apropiado de educación, uno obtendría el cerebro adulto. Presumiblemente, el cerebro de un niño es algo así como un cuaderno que uno compra en una tienda. Un mecanismo más bien

simple, con muchas hojas en blanco. (Mecanismo y escritura son casi sinónimos para nuestro punto de vista.) Nuestra esperanza es que haya tan poco mecanismo en el cerebro del niño, que algo así pueda ser programado fácilmente. Podemos asumir en una primera aproximación que la cantidad de trabajo en la educación sería muy similar a la de un niño humano.

La idea presente de fondo es que la maquina podría ser sometida a enseñanza y con ello aprender y adaptarse a las situaciones. Sin embargo, no ahondare mas en este punto desde la perspectiva de Turing, pues él está pensando en el criterio lingüístico, en cambio yo apelo a otro punto, si es que la maquina puede aprender tiene, con el fin de adquirir una capacidad de adaptación, debió ser programada con alguna finalidad específica, del mismo modo en que las emociones tienen la finalidad de orientar a la mente para regular la química del cuerpo, por tanto si es que la maquina es capaz de tomar conciencia de algún problema dentro de su instanciación corporal (atravesarse la mano por ejemplo), y para ello tiene un cambio en su actitud corporal (mirar y analizar la “herida”) podríamos decir que la maquina tiene un cierto tipo de emoción en ese minuto, luego, si a partir de esa toma de conciencia la maquina consigue un aprendizaje ella luego cambiara su conducta, y su programa relacionara el hecho con un posible daño o peligro manifestando algo similar al miedo. Es cierto que lo que he propuesto no es una emoción propiamente dicha, pero es imposible realmente entender cómo se manifestaría una emoción en una maquina (si es que una maquina pudiese tener una emoción similar a un humano), pero en base a que ambos procesos tienen una finalidad común (mantener la “salud” del cuerpo), se presentan por primera vez en base a un input exterior y requieren de procesos materiales creo posible considerar estas como emociones.

El método emocional

Si es que lo anterior me permite tratar las emociones como un criterio, ahora falta el saber cómo es que se puede constatar ese criterio, a esto lo llamare “método”, en el caso del criterio lingüístico el método para constatar ese criterio es ser un hablante competente y “conversar” con aquello a lo cual se pretende poner a prueba. Si es que este se encuentra a un nivel de conversación suficiente, que demuestre entender lo que se diga y darse a entender debería, por tanto, estar dotado de estados mentales. Sin embargo, las emociones son personales, y a diferencia del lenguaje que puede llegar a generar una convención mediante la cual explicar los pensamientos, las emociones se asemejan más a los *qualia*, por tanto, un acercamiento

directo es imposible. Ya he hablado de las neuronas espejo, pero estas no acreditan más que “algo” se percibe nada acredita que aquello que se “mueve” sea una emoción o algún otro tipo de espasmo muscular. Por tanto, otro criterio debe ser propuesto. Podría asumirse nuevamente al lenguaje como método, en otras palabras, saber que el otro experimenta una emoción en la medida en que él lo manifiesta, pero esto implicaría que un robot que realmente no tiene nada similar a una emoción pudiese engañar a una persona hablante, por tanto, creo que este no es el mejor método posible.

Pero en analogía con el criterio lingüístico, que se ocupa a sí mismo como método, conviene tratar a la emocionalidad, ya he dicho en distintos momentos que llegada a la adultez difícilmente existe algo en el mundo que no se comporte como un EEC, sin embargo, nada impide que en la niñez los seres humanos consideren como un EEC, a los otros seres humanos, es más, parece probable que las actitudes de la madre sean el primer EEC que experimentan las personas, siendo estas con el tiempo capaces de entender las respuestas que generan en sí mismos las distintas actitudes maternas, creando el primer lazo afectivo del sujeto. por tanto, la experiencia le debería posibilitar “emular” lazos afectivos con seres o individuos con los que no exista parentesco, manifestando distintas emociones sociales con ellos, tomándose a sí mismos como referentes y las actitudes ajenas como aquello a probar.

Con este análisis somero puedo postular mi teoría, quedando planteada de la siguiente manera: el individuo mediante la maduración neurológica es capaz de manifestar emociones más complejas y en base a sus experiencias encontrar secuencias entre las emociones que va experimentando a lo largo de su vida hasta identificar patrones; ya llegada cierta edad estos patrones son fácilmente reconocibles y toman como EEC conductas ajenas, de tal modo que al captar ciertas actitudes él sujeto es capaz de predecir la conducta que seguirá el otro y puede “contrastar” al tiempo si es que su teoría se condice con lo que él está experimentando a un nivel emocional, de tal modo que al llegar a una emoción específica él asume que esta frente a otro ser pensante. por tanto, el sujeto al sentir ciertas emociones “sabe” que se encuentra frente a un ser pensante y emocional.

Aun así, no he dado una explicación sobre cómo es que esta toma conciencia de las emociones ajenas, creo que en esto juegan dos factores particulares la “empatía” y las “emociones sociales”.

Ya he explicado que la empatía es la emoción que se experimenta frente a la emoción de otro, y las emociones sociales son aquellas emociones que versan sobre la respuesta que se experimenta frente a la actitud de otro y ese otro mismo, pero cuando ambas se interconectan pueden generar un sistema, por una parte, la empatía tiene una fuerte influencia en las emociones sociales, el desdén, por ejemplo interfiere en la capacidad de las personas para poner sus emociones en concordancia, por tanto la capacidad de “conectar con el otro” depende en mayor medida de que es lo que siente el sujeto mismo. Aun así, es necesaria para la aparición de una emoción social la aparición previa de otros procesos homeostáticos entre ellos otras emociones. por tanto, las emociones primarias que se experimentan en algunos casos generan emociones sociales que a su vez producen las capacidades empáticas, siempre y cuando en conjunto con las emociones primarias se genere la adjudicación de estados mentales (la detección de aquel EEC específico ya mencionado).

Por tanto, la capacidad de mindreading sería en el fondo la sistematización de la emocionalidad de la persona más la toma de conciencia de esta emocionalidad, que al paso del tiempo se perfeccionaría en tanto mayor experiencia emocional se tiene. Una concepción como esta no implica de ningún modo que el sujeto “lector de mentes” tenga realmente algún tipo de certeza, de hecho, es muy sencillo engañarle, pues su método son sus emociones, que respectan solo a él, en cuyo caso el mindreading sería una “apuesta” sobre los estados internos de otro, en función a patrones que se suelen repetir entre los distintos individuos.

una teoría como esta resulta poco plausible sin contrastación empírica, si bien no me es posible exponer un experimento hecho para esta misma, puedo servirme de la tesis de Trinidad Maldonado, quien estudio la relación entre la atribución de estados mentales y la empatía en conjunto con el lenguaje no verbal. El método que utilizo en su trabajo de tesis fue el uso de películas mudas y como es que se realizaban las atribuciones de estados metales frente a los personajes de estas. Sin embargo, su simpatía con la teoría de teorías resulta en una dificultad para mi propia teoría aquí expuesta por lo cual no puede ser tomado como conclusivo.

Una de las pruebas más interesantes fue la proyección de dos videos, el primero tenia contenido verbal y el segundo no; la prueba fue sometida a un grupo de estudio que comprendía niños entre los 3 a los 13 años de edad, y se les preguntaba cuál era el desarrollo

de la historia y cuales emociones experimentaban los personajes, a continuación, la siguiente figura expresa los resultados en tanto a porcentaje de éxito de los participantes:

Fig. 6

	N	Rango Promedio		N	Rango Promedio
3-4 años	7	6.00	3-4 años	7	5.21
6-7 años	7	9.00	6-7 años	7	10.00
13 años	5	17.00	13 años	5	16.70

La primera tabla respecta a la adjudicación de estados mentales teniendo lenguaje verbal mientras que la segunda son las películas mudas. De estas se deducen dos puntos, el primero es que la habilidad aumenta acorde a la edad. El segundo punto es una muestra de que se suele fallar más sin un lenguaje verbal de por medio; sin embargo, esto no desecha mi punto, si se necesitara un lenguaje verbal en realidad la diferencia entre una y otra debiese ser mucho más profunda, pero en la tabla claramente no es así y, en segundo lugar, no se puede obviar que el lenguaje verbal también es un EEC, por lo cual en el segundo caso se “jugaría” con menor información, por lo cual creo que es motivo suficiente para validar una investigación futura, pero no es conclusivo de ningún modo.

En el caso del experimento mental ya tratado mi hipótesis marcaría una clara tendencia a la adjudicación de estados mentales de la máquina, ya que esta al manifestar cualquier cosa similar a una emoción (como sonreír o sorprenderse) debiese manifestar en el espectador una emoción, que generalmente se manifiesta ante seres pensantes por lo cual se le adjudicaría algún estado mental. Y aunque alguien pudiese argumentar que esto no es prueba decisiva sobre si es que esta (el autómata) tiene o no estados mentales, sería obviar que la adjudicación ha sido hecha de todas formas, y que realmente ningún argumento debiese cambiar realmente el convencimiento de esto, de igual manera que resulta casi imposible convencer a una persona cristiana de la inexistencia de dios mediante argumentos racionales.

Conclusiones

Según lo tratado, es posible considerar a las emociones como uno de los mayores factores para la mente en la medida que sin ellas, un cuerpo complejo como el ser humano no tendría algún método para regularse a sí mismo, en segundo lugar, las emociones sociales son fundamentales a la hora de regular las interacciones sociales del individuo, finalmente son una parte constitutiva de la mente. El problema empieza con la adjudicación de estados mentales; soy consciente de que no he brindado elementos suficientes para decir que las emociones son primordiales para el mindreading, sin embargo, he tratado como es que las emociones personales posibilitarán la comprensión de las emociones ajenas en la medida de que exista algún nexo afectivo, pero no a un nivel definitivo. Para despejar esta duda es necesaria la creación de algún experimento empírico, el primer experimento que generaría es alguno en el cual se someta al sujeto a la prueba de la falsa creencia conectado a una máquina capaz de escanear el cerebro y los sectores implicados en la emocionalidad, y el segundo sería un estudio sobre la progresión de esta capacidad con personas que tienen diagnosticado algún trastorno que le impida sentir emociones sociales, popularmente llamados psicópatas.

Además, creo que se abren nuevas alternativas y exigencias para la inteligencia artificial, dándole un mayor énfasis a la instanciación material de esta, de tal modo que la tarea de los programadores sería mucho más extensa con tal de poder crear una inteligencia artificial, pero por otra parte, el que la propia máquina se preocupe de sí misma ahorraría en gran medida los problemas técnicos que esta pudiese tener, y si bien ya existen máquinas que llevan a cabo procesos similares, el complejizar estos mecanismos podría conllevar grandes mejoras para el avance y aprendizaje de estas.

El otro punto que quiero tomar es el análisis sobre el videoclip, en un principio tenía la intención de realizar un análisis más profundo sobre este, sin embargo, el material es muy poco como para ello, y junto con el poco espacio del que dispongo me ha sido imposible. Sin embargo, en algún futuro trabajo pretendo analizar en conjunto el video musical de la misma banda llamado "*Am i happy*", video que fue utilizado como materia promocional del disco en el cual se expande la historia. Lamentablemente este video cuenta con diálogos, por lo cual el análisis emocional sería más complejo de realizar.

Finalmente, estoy consciente de que ha faltado un análisis y contraste contra el criterio lingüístico, puesto que generalmente se asume una relación entre el mindreading y la aparición de las capacidades lingüísticas en los seres humanos. El motivo por el cual no he trabajado este punto, ha sido una decisión complicada de tomar, pero lo he dejado de lado por cuestiones de espacio, ya que si hubiese trabajado este punto, habría tenido que sacrificar el espacio para tratar la emocionalidad en profundidad, sin embargo, para futuros trabajos esta será una de las prioridades a trabajar en conjunto con el nexo entre las capacidades lingüísticas y la emocionalidad.

Bibliografía

- Beorlegui, C. (2007). Filosofía de la mente: Visión panorámica y situación actual. *Realidad: Revista de Ciencias Sociales y Humanidades*, 121-160.
- Blue October. (22 de Febrero de 2018). *Youtube*. Obtenido de Blueoctoberofficial: <https://www.youtube.com/watch?v=FwZvuV79jeY>
- Damasio, A. (1996). *El error de Descartes*. Santiago de Chile: Andres Bello.
- Damasio, A. (2005). *En Busca de Spinoza*. Barcelona: Drakontos.
- Damasio, A. (23 de abril de 2008). El cerebro, teatro de las emociones. (E. Punset, Entrevistador)
- Delgado, J. (2011). Tres debates fundamentales en el campo de la teoría de la mente: Aspectos teóricos y metodológicos. *Acta colombiana de Psicología*, 109-118.
- Delgado, J. (2018). Los niños y la comprensión del mundo mental de los otros: Teorías explicativas. *Psicogente*, 518- 544.
- Descartes, R. (2001). *Discurso del método*. Madrid: Mestas.
- Descartes, R. (2011). *Meditaciones metafísicas*. (G. Grañó Ferrer, Trad.) Madrid: Alianza.
- Descartes, R. (2011). *Meditaciones metafísicas*. Madrid: Alianza.
- Gomila, A. (2002). La perspectiva de segunda persona de la atribución mental. *Azafea*, 123-138.
- González, R. (2007). El Test de Turing: Dos mitos, un dogma. *Katholieke Universiteit Leuven*, 37-53.
- González, R. (2016). Dos criterios para la presencia de estados mentales: Descartes y Turing. *Cinta moebio*, 159-171.
- Gopnik, A., & Wellman, H. M. (1992). Why the Child's Theory of Mind Really - Is a Theory. *Mind and Language*, VII, 145-171.
- Martinez, M. (2011). Intersubjetividad y Teoría de la Mente. *Psicología del Desarrollo*; , 9-28.
- Martínez, M. (2011). Intersubjetividad y Teoría de la Mente. *Psicología del Desarrollo*, 9-28.
- Medina, N. (2008). La ciencia cognitiva y el estudio de la mente. *IIPSI*, 183-198.
- Oberman, L. M., McCleery, J. P., Ramachandran, V. S., & Pineda, J. A. (2007). EEG evidence for mirror neuron activity during the observation of human and robot actions: Toward an analysis of the human qualities of interactive robots. *Neurocomputing*, LXX, 2194-2203.
- Ryle, G. (2005). *El concepto de lo mental*. Barcelona: Paidós.
- Searle, J. (2006). *La mente una breve introducción*. Bogotá: Grupo editorial Norma.
- Tellez-Vargas, J. (2006). Teoría de la mente: evolución, ontogenia, neurobiología y psicopatología. *Avances en Psiquiatría Biológica*, 6-27.

Turing, A. (1950). Maquinaria computacional e Inteligencia.

Zabala, M. L., Richard's, M. M., Breccia, F., & López, M. (2018). Relaciones entre empatía y teoría de la mente en niños y adolescentes. *Pensamiento Psicológico, XVI*(2), 47-57.

Zilber, A. (2017). Teorías acerca de la Teoría de la Mente. El rol de los procesos cognitivos y emocionales. *Neuropsicología Latinoamericana, 1-12*.