



Universidad de Chile

Facultad de Filosofía y Humanidades

Departamento de Filosofía

Corporización

La inteligencia entendida desde la conjunción percepción
– acción en la robótica de Rodney Brooks

Informe Final Seminario de Grado: “Ciencias Cognitivas”, para optar al
grado de Licenciado en Filosofía

Autor:

Javier Godoy Escobedo

Profesores Guías:

Guido Vallejos

Manuel Rodríguez

Santiago, Chile

Marzo 2013

Agradecimientos

Por el motivo de la entrega de esta tesis, quiero agradecer a aquellas personas que me han brindado su apoyo, ya sea de manera afectiva o intelectual, en el proceso de creación de este trabajo. Quiero agradecer a mi familia y amigos por su constante aliento y motivación. Sin su intensa energía no creo que esta tesis se pudiera haber llevado a cabo. Quiero agradecer también a mis compañeros de universidad, quienes han motivado en mí el análisis filosófico de los temas tratados aquí. En este respecto, quiero agradecer especialmente a Darío Oyanadel, Aníbal Osorno y Pablo Contreras, ya que, sin las muchas conversaciones sostenidas con ellos respecto a estos temas, mi comprensión acerca de los mismos sería totalmente distinta. Agradezco infinitamente al grupo de estudios de filosofía analítica de la Universidad de Chile (Gefauch), ya que su gente, los temas tratados, las exposiciones y la experiencia como parte de él, me han brindado la seguridad y los conocimientos imprescindibles para llevar a cabo esta tarea. Mis agradecimientos al profesor Rodrigo González, cuya enseñanza me motivó a desarrollar el carácter necesario para este tipo de filosofía. Agradezco también al profesor Cristian Soto por introducirme en los estudios de la filosofía analítica. Sin su guía el camino tomado por mí en la filosofía hubiera sido, seguramente, muy distinto. Finalmente, quiero agradecer a mis profesores guías de esta tesis Manuel Rodríguez y Guido Vallejos. Agradezco al profesor Rodríguez su infinita paciencia conmigo, su amabilidad y las correcciones que le hizo a este trabajo. Por sobre todo, le agradezco al profesor Vallejos por su estoica confianza en mí además de que, sin su erudición y enseñanza, mi camino a través de la filosofía analítica y las ciencias cognitivas no hubiese sido posible. Por lo mismo, quiero dedicar el presente trabajo al profesor Guido Vallejos, por haberme enseñado el valor de pensar menos y mirar más.

Tabla de Contenidos

Introducción 8

Capítulo Uno: Sentando las bases para la investigación 10

- 1.1 – Algunas nociones acerca de lo inteligente 10**
- 1.2 – ¿De qué manera una situación puede depender de un organismo? 13**
- 1.3 – Deshumanización de la inteligencia 15**
- 1.4 – Criterios de *concordancia* y *diversidad* 18**
- 1.5 – Resumen: Lo mental, lo cognitivo y lo inteligente 20**

Capítulo Dos – El Paradigma *Cognitivista* 22

- 2.1 – El desarrollo y las implicancias del paradigma clásico 22**
 - 2.1.1 – Los dos caminos propuestos por Turing 22*
 - 2.1.2 – Hipótesis de sistema de símbolos físicos 26*
 - 2.1.3 – Problema del dinamismo 30*
 - 2.1.4 – Problema de la relevancia 32*
 - 2.1.5 – Evaluación 34*
- 2.2 – Conexionismo 35**
 - 2.2.1 – Sistema conexionista 36*
 - 2.2.2 – Nivel conceptual v/s nivel sub-conceptual 38*
 - 2.2.3 – Soluciones al problema del dinamismo y al problema de la relevancia 39*
 - 2.2.4 – Evaluación 40*

Capítulo 3 - Post-cognitvismo: Cognición Corporizada 44

- 3.1 – Consideraciones Preliminares 44**
 - 3.1.1 – La mitología *cognitivista* 44*
 - 3.1.2 – Percepción directa y ambiente 46*

3.1.3 – Operadores de máquina 48

3.1.4 – Estableciendo bases para una inteligencia corporizada: Concordancia y Diversidad 50

3.2 - La robótica corporizada de Rodney Brooks 52

3.2.1 – Una conclusión inesperada y una hipótesis radical 52

3.2.2 – Representaciones 54

3.2.3 – Cuatro aspectos fundamentales 55

3.2.4 – Inteligencia y cognición 61

3.2.5 – Criaturas 62

3.2.6 – Hipótesis del enraizamiento físico 63

3.2.7 – Arquitectura de la subsunción 65

3.2.8 – Dos instanciaciones de la metodología 68

Conclusión 70

Lista de Referencias 73

Resumen

Pocos saben a qué se refieren, de manera concreta, cuando hablan de inteligencia. Por inteligencia se puede entender el comportamiento de los animales, de las máquinas y los seres humanos según quién se exprese. Hay muchos, por ejemplo, que defenderían que la inteligencia es un rasgo particular de los seres humanos y no es propio de ningún otro organismo. Pero la verdad es que no existen buenas razones *científicas* para considerar esto último cierto. Es posible que la naturaleza de la inteligencia sea transversal a todos los organismos y no sólo característica del ser humano. Los comportamientos inteligentes en general cumplen con dos criterios: actúan en *concordancia* con las leyes sociales y físicas del ambiente y son *diversos* para responder a las circunstancias que se presentan. Esta manera de ver la inteligencia nos permite restarle relevancia en la explicación de la inteligencia a los mecanismos cognitivos que pudieran determinarla, y darle más importancia, en aquella misma explicación, a la relación de ese comportamiento con el ambiente.

El enfoque clásico en ciencias cognitivas e inteligencia artificial ha mostrado no ser del todo satisfactorio para la comprensión del concepto mismo de inteligencia. Los principios que funcionan bajo la creación de máquinas basadas en la arquitectura clásica han mostrado tener dos problemas que parecen insolubles para el paradigma clásico: un organismo que funciona bajo la manipulación simbólica no puede comportarse de manera *dinámica* con el ambiente y no puede obtener la información que le resulta *relevante* para sus propósitos. El conexionismo no logra ser una alternativa a este acercamiento, ya que presenta, en última instancia, los mismos problemas que el enfoque ortodoxo.

Por eso, debemos buscar una nueva explicación para los fenómenos de la inteligencia y la conducta que nos permita también la creación de robots capaces de sortear los problemas de la inteligencia artificial clásica y los cuales tengan posibilidades de ser inteligentes a la manera que lo son los animales y las personas. Este enfoque nace de la conjunción de los procesos de percepción y acción, dejando de lado los procesos

abstractos que acarrear los problemas mencionados. Una manera de llevar a cabo esto, es la metodología que ha trabajado Rodney Brooks, creando criaturas capaces de comportarse de manera dinámica con el ambiente, permitiéndonos así acercarnos a la comprensión de la inteligencia.

“Cuando levanto mi brazo voluntariamente,
no me estoy sirviendo de un medio para producir el movimiento.
Tampoco mi deseo es un medio semejante.
El querer, si no es una especie de deseo,
debe ser el actuar mismo.” (Wittgenstein, 2009, p. 160)

Introducción

La inteligencia es un fenómeno que nos permite establecer muchas distinciones posibles. Algunos creen que la inteligencia separa a los organismos de intelecto superior como los seres humanos de los demás (no - humanos). Otros piensan que la inteligencia más bien separa a aquellos seres que exhiben comportamiento en vista de sus necesidades (otra manera de decir esto es que son *animados*) de los que no poseen comportamiento alguno (objetos inanimados). Me considero personalmente dentro del segundo grupo. Sin embargo, para llegar a defender esta posición hay que llevar a cabo una suerte de limpieza conceptual del campo de investigación. Si se quiere comprender cuáles son las razones que tenemos para otorgar inteligencia a ciertos organismos, debemos inspeccionar el origen de nuestra terminología así también como las bases de nuestro *sentido común* al respecto. Ahora, no pretendo defender una posición que se base en el sentido común o en meras intuiciones; todo lo contrario, nuestra actitud con nuestras creencias más profundas debe ser crítica en el afán de tener la fuerza suficiente para desecharlas si es que resultan ser falsas. Recordemos que es altamente de sentido común pensar, al mirar el cielo, que es el Sol el que gira alrededor de la Tierra. No debemos dejarnos engañar: sólo la constante experimentación científica nos mostrará que tan cerca estamos del entendimiento del fenómeno de la inteligencia.

La cognición corporizada es aquel planteamiento en ciencias cognitivas que afirma que cualquier cosa que sea la inteligencia o la cognición, debe darse en un organismo que posea un cuerpo, es decir, sensores que le permitan extraer información del mundo físico así como actuadores que le permitan comportarse de acuerdo a sus necesidades. Una de las razones para esto salta a la vista: los seres humanos poseen cuerpos, y son el modelo de la inteligencia por excelencia. Programar una máquina para jugar ajedrez a través de procesos abstractos no es tarea difícil, pero como Gomila y Calvo han dicho (Calvo & Gomila, 2008, pp. 1 - 2) muy diferente es programar una máquina para jugar pool. El pool es un juego que necesita de la dinámica de un cuerpo, es una actividad *corporizada*. Así lo son la mayoría de las

actividades a las que se enfrentan las personas y los animales, no son problemas abstractos.

A lo largo de este trabajo intentaré ilustrar como nuestra manera de entender la cognición y la inteligencia han sido determinadas por la creciente influencia de los computadores en los últimos 60 años. Muchas de las concepciones que manejamos son más bien de tipo computacional y no están basadas en la realidad que podemos observar del mundo. La investigación que ahora se expondrá se divide en tres partes, y las últimas dos a su vez, en dos más. El *primer capítulo* intenta generar problemática respecto a la terminología de lo inteligente, la conducta y lo cognitivo. Esta parte es a lo que comúnmente se le llama *bombardeo de intuiciones*. Su objetivo es revelar los primeros fundamentos de lo que entendemos por inteligencia así como mostrar hacia donde se debiera dirigir una investigación acerca de aquello en las ciencias cognitivas y en la inteligencia artificial (desde ahora IA). El *segundo capítulo* es un análisis de dos propuestas metodológicas cognitivistas. La primera es la que recibe el nombre de *cognitivism clásico* u *ortodoxo*, y se ilustra a través de la *hipótesis de sistema de símbolos físicos* de Newell y Simon fundamentalmente. El segundo enfoque es el *cognitivism conexionista* que es abordado desde las descripciones hechas por John Tienson (1995), Paul Smolensky (1988) y David Rumelhart (1997). El *tercer y último capítulo* también se compone de dos secciones. La primera es una introducción a las intenciones de los enfoques alternativos. Allí se ilustra el objetivo de éstos de querer sentar un nuevo paradigma para la explicación de la conducta en ciencias cognitivas y en IA. La segunda sección consiste en la exposición de la filosofía e ingeniería robótica del creador de robots Rodney Brooks. Allí se analiza la metodología que ha llevado a Brooks a sostener un enfoque corporizado con respecto al entendimiento de la cognición y la inteligencia, el cual, plantea que la cognición es el acoplamiento de los procesos de percepción y acción de un organismo.

Capítulo Uno – Sentando las bases para la investigación

1.1 – Algunas nociones acerca de “lo inteligente”

Hablar acerca la inteligencia es una tarea bastante difícil. Las concepciones que en general se manejan acerca de ésta suelen estar arraigadas muy fuertemente en la tenacidad¹ de las personas. Sin embargo, nadie estaría dispuesto a exponer con mucha seguridad dónde residen las condiciones para que la inteligencia acontezca, y, estarían aún menos seguros, de decir por qué es que aparece en esos casos y no en otros. Comúnmente nos sentimos bastante seguros de considerar algunos fenómenos como muestras de inteligencia y, en muchos casos, estos fenómenos parecieran ser causa de naturalezas distintas. La máquina que derrotó a Kasparov en una partida de ajedrez en el año 1996 (*Deep Blue*) puede ser considerada, desde alguna perspectiva, inteligente. También puede ser considerada inteligente la manera en que un chita prepara, acecha y caza a su presa, o, por ejemplo, recuerdo que cuando mi sobrino comenzó a caminar a temprana edad pensé que aquello denotaba en él una inteligencia superior a la común. Si todos los comportamientos descritos pueden ser tildados en la mayoría de las ocasiones de *inteligentes*, propongo dos hipótesis que pueden explicar esto:

H1) Tenemos una muy ambigua y equívoca concepción acerca de lo inteligente, ya que los fenómenos descritos son tan diversos por naturaleza que no pueden compartir los tres algo así como “la propiedad de lo inteligente”.

H2) Tenemos una muy ambigua pero sensata concepción de inteligencia, lo que permite creer que si los tres fenómenos conductuales pueden ser tildados de inteligentes, los tres comparten “la propiedad de lo inteligente”.²

¹ Charles Peirce habla acerca de los métodos de cómo las personas mantienen sus creencias (Peirce, Ch., S. (1992) “The fixation of belief” en *The Essential Peirce, Selected Philosophical Writings, Volume 1* Ed. Nathan Houser and Christian J. W. Kloesel. Indiana Indianapolis: University Press). La tenacidad es uno de esos métodos.

² Aquí cuando digo *propiedad de lo inteligente* no me refiero de ninguna manera a que haya algo así como *lo que hace a los organismos esencialmente inteligentes* de manera que la correspondencia de esa propiedad con el concepto sea “Si denota la propiedad A, entonces es inteligente”. Lo que intento simplificar con este lenguaje es el hecho de que, si bien, no existe algo así como *lo inteligente* (punto que será abordado más adelante en este mismo capítulo), sí hay similitudes entre las cosas que llamamos

Si aceptamos la primera o la segunda explicación ya no dependerá de nuestras intuiciones, sino de lo que el mundo y la experimentación en él puedan decirnos. La ciencia cognitiva clásica tiene una larga historia con respecto a la investigación de la pregunta por la inteligencia. Ésta pregunta engloba también a otros términos propios de la ciencia cognitiva, como *cognición y mente*, entre otros. Dada la fuerte conexión que existe entre esta terminología, creo que dar una explicación acabada de la inteligencia es también, en algún grado, poder dar cuenta de términos como los anteriores.

Entonces, ¿cómo podemos empezar a abordar una investigación con respecto a los fenómenos de la inteligencia? Primero hay que especificar, de manera muy parcial, cuáles deben ser los propósitos de nuestra especificación conceptual. Debido a que este escrito intenta abordar esta problemática desde la discusión que se origina en las ciencias cognitivas, me parece que lo inteligente está primeramente definido por el *comportamiento de los organismos biológicos*³. Es menester preguntarse ahora si cualquier conducta ejecutada implica necesariamente inteligencia. Si yo me encuentro ante un asaltante ¿sería inteligente de mi parte practicar mis recientemente aprendidas técnicas de judo con él?, ¿Lo sería por otra parte, orinarme en los pantalones? Veámoslo desde otra perspectiva. Si colocamos ante el tablero de ajedrez a alguien que al comenzar el juego se comporta de manera que dice: “Lo siento, pero no sé jugar ajedrez”, ¿sería en algún sentido menos inteligente que *Deep Blue*? Si es concedida la hipótesis de que para que algo sea llamado inteligente debe ser propio de la conducta de un organismo biológico, podemos entender por qué una persona que sólo atina a orinarse ante un asaltante no deja de ser un organismo inteligente, y así mismo por qué a pesar de que *Deep Blue* se comporta como un ajedrecista profesional no es *necesariamente* un organismo inteligente. Más claramente, una manera de saber si un organismo es o no inteligente es esperar a que se comporte y preguntarse: (i) ¿Se

inteligentes de la misma manera que Wittgenstein habla que hay concepciones como “juego” que no tienen una propiedad esencial pero sí comparten ciertos *parecidos de familia*. (Wittgenstein, 1999)

³ Esta consideración no es una petición de principio. A lo largo de este capítulo y también del capítulo número tres se argumentará porqué hay razones suficientes para abordar la investigación de la inteligencia en ciencias cognitivas desde este punto.

comportaría así ante aquella situación un organismo biológico competente? En este respecto, competencia significa que el organismo debe contar con una integridad física plena, además de tener todas las posibilidades de conducta, que su naturaleza permita, a disposición. Este método, sin embargo, no está exento de problemas. Si somos demasiado rigurosos y fieles a él nos encontraríamos ante una gran disyuntiva ante el testeo de organismos, digamos, en la misma “prueba del ajedrez”. Se vio antes que existen por lo menos dos tipos de organismos biológicos que especificarían dos conductas diferentes ante, al parecer, la misma situación: el jugador de ajedrez y el no jugador de ajedrez. Bajo esta realidad nos damos cuenta que el método de aplicar (i) resulta banal, puesto que en el testeo experimental podríamos colocar tanto una máquina experta en ajedrez como un pañuelo de seda y ambos darían una respuesta afirmativa a (i): la máquina se comporta como un organismo biológico jugador de ajedrez al jugar ajedrez y el pañuelo se comporta como un organismo biológico no jugador de ajedrez al no jugar ajedrez. Hay dos consideraciones que me gustaría hacer ante este respecto.

(C1) Para mejorar nuestra metodología en la búsqueda de inteligencia en términos de la inteligencia de organismos biológicos, debemos preguntarnos si acaso, en los casos señalados (de la máquina de ajedrez y el pañuelo), los organismos *realmente* se comportan como organismos biológicos. Cuando digo *realmente* quiero señalar si acaso los organismos determinan⁴ su conducta *de la misma manera* que lo haría un animal o persona.

(C2) Cuando decimos que la máquina de ajedrez y el pañuelo se encuentran en la misma situación que los organismos biológicos “ajedrecista” y “no ajedrecista” nos estamos, quizás, apresurando. ¿Es la situación independiente o dependiente del

⁴ El cómo un organismo biológico determina su conducta es algo que se explicará cuando se planteen los criterios de concordancia y diversidad para la identificación de conductas inteligentes en la sección 4 de este capítulo y en la sección 1.4 del tercer capítulo. En aquellas secciones se postula que las conductas determinadas por los organismos que llamamos inteligentes ocurren siempre en concordancia con las leyes físicas y sociales, además de ser conductas diversas que responden a un ambiente en constante variación.

organismo que la instancia? Este respecto será el abordado ahora y su análisis propone explicar (C1) y (C2).

1.2 – ¿De qué manera una situación puede depender de un organismo?

Es fácil, a primera vista, reconocer que existe algo así como la conducta y el organismo que la ejecuta. En vista de querer resaltar los aspectos relevantes que determinan la primera, hemos también de hacer referencia al *ambiente*.

Para una explicación más clara, imaginemos que un amigo me regala un libro escrito completamente en chino. La verdad es que yo no sé nada de aquella lengua milenaria, por lo que mis posibles acciones con respecto al libro serán o dejarlo en la estantería como adorno, usarlo para mantener abierta la ventana de mi pieza, o con la ayuda de un diccionario y algunos libros sobre gramática, comenzar mi duro camino para aprender a leer el chino y así poder leer el libro. En cambio, si mi amigo le hubiera regalado el mismo libro a un asiduo lector de literatura china su más predecible acción con respecto al libro hubiera sido leerlo (todo esto dentro del ámbito de la conjetura, porque también puede ser el caso de que ya lo hubiese leído y no quisiese hacerlo otra vez, lo que también es un comportamiento particular de una persona que habla chino). Salta ahora a la vista que los comportamientos descritos para cada caso son conductas inteligentes. Si es así, podemos explicar este tipo de conductas diciendo que cada vez que el organismo se encuentra frente a alguna situación (se encuentra en un ambiente determinado) tiene un número de conductas específicas bajo las cuales proceder. Éstas son determinadas por las naturalezas tanto del agente como del ambiente. Ante esto ¿por qué no decir simplemente que los organismos son diferentes y que la información que ya poseen determina el contenido de sus percepciones y así la descripción del ambiente es irrelevante? Debemos escoger dónde vamos a ubicar el peso de nuestra explicación⁵. Uno de los postulados que sostendrá esta tesis es que

⁵ “El peso de la explicación” es ampliamente tratado en (Rowlands 2004). En líneas generales, uno debe decidir si la información que permite la ejecución de la conducta de un organismo se encuentra en

ubicar la explicación de aquellos procesos que se llaman cognitivos tanto fuera como dentro del organismo en lugar de sólo dentro, es beneficioso para una mejor explicación de la conducta⁶. No hay razones por ahora para aceptar una explicación de esta naturaleza, pero tampoco es difícil conceder que podemos dar una explicación a los fenómenos conductuales como el del no jugador de ajedrez sin dejar de concederle al organismo su inteligencia. Creo que es posible postular que los organismos *instancian* ciertas posibilidades conductuales de manera diversa, no habiendo una correspondencia *correcta* o de uno a uno entre una conducta y la instanciación de la inteligencia. Si hay dos organismos diferentes, no importa que uno maneje las variables ambientales de manera que se den de manera exactamente igual para ambos. El ambiente será siempre diferente ante un organismo diferente; el ambiente instancia *posibilidades de acción* en conjunto con el organismo.⁷

Es un pensamiento común en filosofía y en psicología que un organismo es inteligente en la medida que puede dar cuenta de una situación que *el experimentador cree relevante* para la verificación de su inteligencia. Sin embargo, esto deja de lado un hecho importante del mundo real, a saber, que lo que determina cuáles conductas son las apropiadas para catalogarlas como inteligentes es el *mundo*. Ahora, el éxito en conseguir objetivos que permiten satisfacer las necesidades del organismo, así como su subsistencia, funciona, en esta explicación, como *sinónimo de inteligencia*. Esto nos lleva a plantear que cualquier conducta que nosotros estimemos pertinente para un organismo es irrelevante en la medida que lo que decide si la conducta es o no pertinente es la reacción del ambiente ante dicha conducta. Una paloma puede no saber jugar ajedrez, pero sí se desenvuelve de manera bastante exitosa al lidiar con las situaciones en las que se encuentra a diario. En este sentido la hipótesis fuerte defendida en este trabajo, propone que *la inteligencia o lo inteligente es esencialmente propio de los organismos biológicos competentes y eso se basa en la manera en que determinan su comportamiento*. Esto se funda en el hecho descrito: el ambiente

mayor parte dentro de él (*internismo*) o fuera de él (*externismo*). La noción de inteligencia defendida en esta tesis es altamente compatible con las visiones *externistas* acerca del fundamento de la conducta.

⁶ Ver Cap. 3, Sec. 1 completa.

⁷ Para una profundización de este punto véase Cap. 3, Sec. 1.2

natural dispone de sus propias condiciones para otorgarle inteligencia a los organismos.

Si queremos conceder una concepción de inteligencia que nos acerque más a la naturaleza de las conductas humanas y animales, deberíamos aceptar que la inteligencia está estrechamente ligada o lo *dinámico* que puede ser el organismo para dar cuenta de las situaciones en la que el mundo lo dispone. La *dinamicidad* es, en pocas palabras, la capacidad de un organismo de tener comportamientos versátiles ante el constante cambio que va sufriendo su ambiente. Al final de este capítulo y en el capítulo 3 sección 2.3 se vuelve a recalcar la importancia de esta propiedad para el entendimiento de los comportamientos inteligentes.

1.3 - Deshumanización de la inteligencia

Hay otra creencia profundamente arraigada, no sólo en ciencia y filosofía sino que también en el sentido común, de que la inteligencia, sea lo que sea eso, se encuentra en mayor grado, o incluso exclusivamente, en los seres humanos. Podemos remontarnos a muchos autores filosóficos en los cuales podemos encontrar, ya sea de manera explícita o implícita, indicios acerca de esta concepción. La remisión más conocida es el filósofo moderno René Descartes quien, por ejemplo, en su *Discurso del Método*, resta a los animales de los dominios de la inteligencia por no poseer lenguaje (Descartes, 1937, p. 5). He aquí otro prejuicio tenaz en la historia de la filosofía: el lenguaje es signo de inteligencia, o más específicamente, el lenguaje es signo de pensamiento, así la línea del argumento sería que: si habla, piensa y si piensa, debe ser inteligente. Hay un chovinismo humanista para nada sutil en psicología con respecto a los roles que juegan nuestras creencias en nuestro comportamiento. Sin embargo, como he estipulado ya, al creer que la inteligencia es propia de los organismos biológicos y no sólo de los humanos, se vuelve más difícil sostener esta premisa de “hablo y/o pienso, luego, soy inteligente”. Al parecer el lenguaje que poseen los animales para comunicarse no poseía las mismas cualidades que el nuestro para Descartes o para Aristóteles (Aristóteles, 2013), porque sea lo que sea eso que

llamamos pensar, es propio de los humanos dada la *composicionalidad*⁸ de nuestro lenguaje. Es aquí donde nace una ligazón conceptual fuerte entre lo que es la *inteligencia* y lo que es la *cognición*, porque es a lo que la psicología llama cognición, es decir, la manipulación de creencias y deseos para la deliberación y la determinación del comportamiento, lo que nos separa de los demás animales. La pregunta que cabe hacerse ahora es ¿y hay buenas razones para creer que lo que denota el comportamiento en los humanos es diferente a lo que denota el comportamiento en los animales? Un psicólogo podría responder a esto diciendo que o bien, las creencias y deseos sí son sólo propias de los seres humanos, y entonces caemos en un *chovinismo humanista acerca de la inteligencia* explícito, o bien, que los animales tienen algo así como creencias y deseos en su mente, lo que determina su conducta igual que en las personas. Cualquiera de estas aseveraciones requiere justificación. Por otro lado, en esta tesis se tomará un camino diferente. Dada la naturaleza de la propuesta acerca de la inteligencia defendida en este trabajo, se podrá ver que, bajo los supuestos expuestos, es sensato adscribirle inteligencia a todo aquello que “nos parezca inteligente”⁹.

Una de las razones que se tienen para no aceptar en principio ninguno de los caminos propuestos por el “sicólogo” es que, plantear que los estados mentales pueden ser causa del comportamiento puede llevarnos (quizás nos lleva obligatoriamente) a una interpretación dualista de la explicación de lo mental. Si se postula una entidad cualquiera, cuyo objetivo es llenar los baches explicativos que dejan las entidades físicas y observables en la teoría, entonces pareciera saltar a la vista que estamos ante una entidad que cubre las necesidades “no físicas” del sistema. Lo que defenderé en esta tesis se basa en la preconcepción de que no tenemos por qué aceptar la existencia de dichas entidades dentro de la explicación de la conducta si estas son por una parte

⁸ La composicionalidad es aquella propiedad que le permite al lenguaje poseer constituyentes *tipo*, por lo tanto poseer una gramática. Una amplia consideración de esta propiedad del lenguaje es hecha por el filósofo Jerry Fodor en la mayor parte de su obra (véase especialmente Fodor, J. A. (2008) “*The language of thought revisited*”. Oxford: Clarendon Press).

⁹ Véase la discusión de porqué plantear una explicación de la inteligencia que haga uso de mecanismos internos no es del todo satisfactoria para nuestro propósitos en el capítulo 2, y la nueva propuesta en cap. 3, sec. 2 completa.

inobservables y por otra parte *insatisfactorias* para la teoría¹⁰. En cambio, los defensores de *teorías alternativas en psicología y ciencias cognitivas*, sostienen que es posible dar cuenta de manera satisfactoria tanto de la conducta así como de la inteligencia sin necesidad de recurrir a la postulación de entidades mentales las cuales no son reducibles del todo a entidades físicas¹¹. Pero dejemos que el presente trabajo justifique esta posición.

Si nos alejamos del racionalismo, es posible interpretar la creencia de que la compleja estructura del lenguaje humano es signo inequívoco de inteligencia es completamente arbitraria. Sólo es defendible en virtud de querer otorgarle a la especie humana una posición aventajada con respecto a las demás especies, pero, no en miras de una tesis consistente con la psicología experimental y/o la robótica. Brooks cree que en general se adscribe inteligencia a aquellas conductas que parecen complejas, y en general se piensa que las conductas de los seres humanos son el pináculo de las conductas complejas (Brooks R., 1991a, p. 19). Sin embargo no es necesario que porque una conducta sea compleja, eso implique algún tipo de procesamiento complejo¹². El ejemplo que ocupa Brooks es *Herbert*, un robot creado por él, el cual está diseñado para tomar latas de soda y llevarlas a su posición de inicio. Desde fuera pareciera que lo que hace Herbert es planear y modelar su conducta, pero lo que hay realmente dentro son muchos procesos simples los cuales ocurren de manera simultánea y carecen de un procesador central. Por otra parte, con respecto a la incidencia posible de *estados mentales* en la causación de la conducta, son conocidos los experimentos del neurocientista Benjamin Libet (Pfeifer & Bongard, 2007, p. 6) en donde se le pide a la gente que mueva su dedo en el momento que ellos estimen conveniente, mientras, los investigadores están a cargo de visualizar la acción neural de los individuos del experimento. El experimento deja entrevisto que la acción neuronal que lleva a una persona a mover deliberadamente un dedo comienza mucho antes de que aquella

¹⁰ Véase cap. 2 secciones 1.3 y 1.4.

¹¹ Esta es la típica discusión del problema mente – cuerpo. Creer que existen dos *sustancias* diferentes las cuales ninguna es reducible explicativamente a la otra es postular también dos naturalezas de explicación diferentes.

¹² Como el sistema de símbolos físicos descrito en el capítulo 2.

persona se lo proponga, es decir, antes de que pudiera poseer algo así como el estado intencional de querer mover el dedo. Esto, es un buen apoyo para querer dejar fuera de la explicación de la conducta o de la inteligencia a los *estados intencionales*, ya que es probable que su participación en la cognición no sea relevante.

1.4 – Criterios de *concordancia* y *diversidad*

Como consecuencia de lo ya expuesto, creo que nos vemos en la necesidad de buscar una nueva explicación para la inteligencia que se aleje de los vicios que arrastra aquella noción tanto en filosofía y ciencia como en el sentido común. Al comienzo de este capítulo dejé entrevisto que nuestro camino era tomar una de las hipótesis que fueron propuestas en un comienzo, siempre en virtud de la experimentación. Todo lo dicho anteriormente tiene como motivo en esta parte apoyar la idea de que no tenemos buenas razones para no concederle inteligencia a las cosas que, de hecho, se consideran comúnmente inteligentes (personas y animales sobre todo) por lo que no deberíamos tener problema de optar por (H2). También en el comienzo del primer capítulo se hizo referencia a que no existe algo así como “la propiedad de lo inteligente” sino, que existen un conjunto de comportamientos los cuales acontecen en el mundo y que son inteligentes en la medida de que son propios de un organismo biológico y “algo más”. Es difícil caracterizar este “algo más”, pareciera que la categoría de comportamientos inteligentes es más bien una categoría disyuntiva, es decir, que los comportamientos A , B , y C sean inteligentes no significa que comparten la propiedad k_1 que es signo de inteligencia, sino que, quizás, que A y B comparten k_1 y que B y C comparten f_1 ambas propiedades suficientes pero no necesarias para la inteligencia. Estas propiedades son instanciadas por los contextos o ambientes en los que se encuentran los organismos. Si el premio nobel de física cae al Amazonas, sus expectativas de supervivencia son estadísticamente menores que las de un nativo de aquella zona. Con este ejemplo también se revela el componente *social* que tiene la inteligencia, ya que lo que es considerado inteligente en la cultura occidental dista mucho de lo que es necesario para sobrevivir en el Amazonas. Esto no quiere decir que el premio nobel de física no sea inteligente, pero sí quiere decir que instancia sólo

una (o algunas) de la(s) propiedad(es) de la inteligencia (digamos k_1) al igual que el nativo del Amazonas (que digamos, instancia f_1).

Rolf Pfeifer y Josh Bongard proponen dos criterios para la identificación de los comportamientos inteligentes:

“A pesar de todas las dificultades de dar una definición concisa de lo que es la inteligencia, y sin tener en cuenta las enormes complejidades implicadas en el concepto mismo, pareciera que cualquier cosa que intuitivamente vemos como inteligente está siempre acompañada de dos características particulares: *concordancia y diversidad*. En resumen, los agentes inteligentes siempre *actúan en concordancia* [*comply*] con las reglas físicas y sociales de su ambiente, y aprovechan esas reglas para producir una conducta *diversa*” (Pfeifer & Bongard, 2007, pag. 34)¹³

Si un organismo se encuentra ante una situación cualquiera, el comportamiento que ejecute deberá estar en concordancia, o, ser ejecutado conforme a las leyes físicas y sociales de su entorno. No hay una especificación explícita con respecto a lo que son estas leyes sociales. Pero, siguiendo con el ejemplo del premio nobel perdido en el Amazonas, las leyes sociales que determinan los comportamientos inteligentes en esta situación son diferentes de las que lo puedan determinar en una reunión con su círculo académico más próximo. Como ya se ha expuesto, la inteligencia parece ser una propiedad *relacional*: determinada por la situación y el organismo de manera bilateral. De esta manera, el premio nobel deberá ejecutar comportamientos que instancien la propiedad f_1 y no k_1 , en vista de que su comportamiento sea considerado inteligente (o en vista de que su comportamiento sea exitoso con respecto a su propósito).

La diversidad de la conducta es también una de las propiedades de la inteligencia según la apreciación de Pfeifer y Bongard. Sobre todo en lo que respecta a situaciones

¹³ Nota: Todas las citas que se encuentren en español y cuyas referencias, al final de este trabajo, se muestran en otro idioma (en todos los casos, inglés), fueron traducidas por el autor de esta tesis: Javier Godoy Escobedo. Para especificar esto, desde ahora al final de cada cita, cuando sea el caso mencionado, se encontrará: (Trad. Autor).

del día a día, nos damos cuenta que cada situación (por ejemplo, ir a comprar pan en la mañana) tiene una diversidad de maneras de ser llevada a cabo. Estas maneras (correr hacia el almacén, caminar hacia el almacén, pagar con monedas de a 100, pagar con monedas de a 500, etc.) hacen que la capacidad del organismo de comportarse se enriquezca y pueda ser determinada en virtud de su eficiencia o de la complicitad que conecte al organismo con ciertos comportamientos. Lo que quiere desmentir el criterio de diversidad es que comportarse de manera inteligente sea de alguna manera *comportarse de manera específica*. Al contrario, al ser el ambiente una constante sucesión de variabilidades, el organismo debe tener los recursos suficientes para poder responder a las emergencias que sucedan durante la tarea programada.

Un recorrido más amplio y profundo en estos criterios y otros se abordará en el capítulo 3, donde se explica con más detalle la caracterización de esta noción de inteligencia para dar lugar a una teoría corporizada en ciencias cognitivas

1.5 - Resumen: Lo mental, lo cognitivo y lo inteligente

Este capítulo ha tenido como objetivo reconsiderar la ambigua y por momentos dogmática idea que tenemos acerca de la *inteligencia*. Es común creer que lo mental, lo cognitivo y lo inteligente son intercambiables *salva veritate* en todas las ocasiones que queremos caracterizar algo como poseedor de alguna de aquellas características. En el transcurso de esta tesis intentaré mostrar que lo que es propiamente inteligente no tiene que ver con lo cognitivo y mucho menos con lo mental. La *deshumanización* de la inteligencia tiene como propósito aquello: proponer que todo aquello que se comporta en concordancia con su ambiente y que posee una diversidad de comportamientos para desenvolverse puede ser llamado inteligente sin reparos. La inteligencia es más bien algo que piensa y atribuye el observador cuando ve la conducta de otro organismo. Es decir, la inteligencia (y como se verá más adelante, también la cognición) ocurre en el ojo del testigo (Brooks, R., 1991a, p. 16 y Moravec, H., 2013).

En el siguiente capítulo se abordará lo que ha sido la investigación de la ciencia cognitiva clásica durante estos últimos 60 años en su búsqueda por una explicación de la inteligencia y la cognición. Me centraré sobre todo en los supuestos en los que se ha basado para plantear sus hipótesis para luego exponer algunos problemas a los que conlleva adoptar el enfoque clásico.

Capítulo Dos – El Paradigma Cognitivista

2.1 – El desarrollo y las implicancias del paradigma clásico

Los inicios de la investigación de las ciencias cognitivas como también del desarrollo de la IA, tienen dos puntos de partida comunes. Estos son la publicación del artículo *Computer Machinery and Intelligence* (1950) por parte de Alan Turing en 1950 y la posterior reunión de investigadores de diferentes áreas que buscaban abordar los problemas de la inteligencia y la cognición en la conocidísima *Conferencia de Dartmouth* que tuvo lugar en el año 1956. Ambos se consideran eventos en los que se comenzaron por primera vez a abordar cuestiones que luego formarían propiamente parte de lo que hoy llamamos *Ciencias Cognitivas*. No es, sin embargo, la intención de este capítulo hacer una reseña histórica acerca del desarrollo de estas disciplinas. La razón por la que menciono estos dos eventos es porque, para explicar la investigación que se ha llevado a cabo por la “buena y anticuada inteligencia artificial”¹⁴ y las posteriores consecuencias de ésta, tomaré como referencia una propuesta algo olvidada hecha por Turing así como también el desarrollo de las hipótesis de Newell y Simon (Newell y Simon, 1985), dos asistentes a la *Conferencia de Dartmouth*.

2.1.1 – Los dos caminos propuestos por Turing

Alan Turing es un referente obligado cuando se habla acerca del desarrollo de la inteligencia artificial, pero más aún cuando se habla de los orígenes de las *máquinas computadoras*. Fue también él uno de los primeros en plantearse de manera seria la pregunta acerca de si las máquinas pueden o podrían eventualmente pensar. Por lo mismo, fue uno de los primeros en poner en discusión nociones como *inteligencia y cognición*.

¹⁴ “BAIA”. Este término fue inventado por John Haugeland en *Artificial Intelligence: The Very Idea* (Cambridge: Harvard Books, MIT Press, 1985), para referirse al paradigma clásico de la ciencia cognitiva.

Para comprender el desarrollo de la noción de inteligencia de Alan Turing, rescataré dos de sus planteamientos centrales desde su texto de 1950. Estos son *el juego de la imitación* (conocido universalmente como *el test de Turing*) y la *máquina de propósito universal* (o *máquina de Turing*). Principalmente, me referiré a las metodologías que podríamos adoptar para poder sortear algo así como el *test de Turing*, a las que Turing hace una importante referencia.

Turing comienza su texto preguntándose si las máquinas podrían de hecho pensar. Él mismo entiende lo ambigua e interpretativa que puede ser esta pregunta (a qué nos referimos con que una máquina piense) por lo que Turing propone un *test* (el juego de la imitación) el cual podría responder la pregunta por la inteligencia de las máquinas. Esto lo hace concordando, quizás, en lo más transversal que poseen las nociones comunes de inteligencia: que los seres humanos la poseen o la exhiben en su comportamiento. En palabras simples, el juego de la imitación consiste, primeramente, en dos sujetos (un hombre y una mujer) quienes se ubican en habitaciones separadas y son interrogados por un tercer sujeto a través de un teletipo. El interrogador debe intentar dilucidar quién es el hombre y quién es la mujer. El objetivo del hombre es que el interrogador falle, es decir, él debe intentar hacerse pasar por una mujer. En cambio, la mujer debe ayudar al interrogador dando respuestas verdaderas a sus preguntas. Podemos ahora reemplazar a los participantes (los interrogados) por un hombre y una máquina, cumpliendo ambos básicamente los mismos roles que los sujetos del primer ejemplo: la máquina debe intentar hacerse pasar por una persona y el hombre debe ayudar al interrogador a que éste haga una correcta distinción entre ambos. De esta manera, la pregunta de si acaso puede una máquina pensar se reduce a si la máquina es capaz de engañar al interrogador y logra hacerse pasar por una persona. Si la respuesta es afirmativa, entonces la pregunta de si la máquina puede pensar o no (signifique lo que signifique eso) es irrelevante, puesto que al poder comportarse la máquina *como una persona* hemos de concederle las mismas cosas que le concedemos a las personas (porque perfectamente mi vecino podría ser una máquina muy efectiva en el juego de la imitación). Este juego es a lo se

llama hoy en día el *Test de Turing* y se ha transformado (quizás inconscientemente) en una especie de meta para los investigadores de la IA y la Ciencia Cognitiva.¹⁵

El juego propuesto por Turing está basado, contrariamente a como se ha seguido su investigación, en un punto de vista casi *conductista*: lo que Turing nos dice a través del juego es que cualquier máquina que se comporte como un humano ante un juez antes preestablecido, podría en teoría pensar y por lo tanto, sería *inteligente*. Turing es abiertamente neutral con respecto a qué metodología es la que se debe seguir para poder sortear con éxito el juego. Lo que nos dice el filósofo y matemático inglés es que lo que debemos conseguir es aquella *emulación* de la mente humana que permita el engaño del interrogador.

Brooks rescatará este punto en *Intelligence without Reason* (1991a, p. 5) a propósito de la siguiente cita del artículo de Turing:

“Podemos esperar que las máquinas eventualmente compitan con hombres en todos los campos puramente intelectuales. Pero ¿cuáles campos son los mejores para empezar? Incluso ésta es una decisión difícil. Mucha gente piensa que una actividad lo bastante abstracta, como jugar ajedrez, sería lo mejor. También puede sostenerse que es mejor proveer a la máquina con los mejores órganos sensores que el dinero pueda comprar, y luego enseñarle a entender y hablar inglés. Este proceso podría seguir la enseñanza normal de un niño. Las cosas serían señaladas y nombradas, y así. Repito que yo no sé cuál es la respuesta correcta, pero creo que ambas aproximaciones deben ser tratadas.” (Turing, 1950, pág. 56) (Trad. Autor)

¹⁵ Esto desde mi punto de vista, el cuál, será justificado a continuación. El test de Turing ha sido desestimado por la misma naturaleza conductista defendida en este capítulo por John Searle (Searle, J. (1980) "Mind, brains and programs". Behavioral and Brain Sciences 3, 417-24) y por Block (Block, N. (1995) "The mind as the software of the brain". Extracted and edited in: J. Heil (ed) Philosophy of Mind: A guide and anthology. Oxford, OUP, pp. 267-274.)

Brooks cree que la ciencia cognitiva clásica decidió optar por el primer camino propuesto por Turing en desmedro del segundo. La solución de problemas abstractos por parte de máquinas *computadoras de símbolos* se volvió el *modus vivendi* de lo que se denomina BAIA. Inclusive aún cuando se han presentado problemas para alcanzar el objetivo de concebir una máquina plenamente inteligente (Cap. 2, secciones 1.3 y 1.4) los cuáles, sin duda alguna, inquietarían al mismo Turing y lo harían diversificar los métodos, el *establishment* se ha mostrado bastante estoico en reconsiderar sus planteamientos.

Tampoco podríamos decir de ninguna manera que Turing fue un precursor de la *corporización*, puesto que las máquinas que él identifica como capaces a la larga de pasar la prueba, son de hecho, *computadoras*. Pero sí podríamos decir que Turing era un cauteloso de la abstracción. Hay una clara intuición en esta parte de su trabajo de que hay elementos que se le podrían escapar a la programación y que, por lo tanto, la investigación podría necesitar que la máquina estuviese *situada* y que fuera a la vez completamente *autónoma*¹⁶ respecto a su aprendizaje.

Que Turing crea que las máquinas que se deban programar para pasar el juego de la imitación deban ser computadoras no es coincidencia. Antes (2004) Turing ya había planteado su manera de concebir una máquina de *propósito general*. Esta máquina (la máquina de Turing), si bien es una idealización, es también la base teórica de lo que son las computadoras¹⁷. Esta máquina consiste simplemente en un aparato que contiene una cinta infinita donde se pueden escribir símbolos. Estos símbolos son leídos y el comportamiento se determina en base a los estados de máquina (combinaciones de símbolos). La cinta es infinita en la medida que corre por la máquina *ad infinitum* siempre proveyendo de un nuevo símbolo a la máquina.

¹⁶ La autonomía y la situacionalidad son características fundamentales para la propuesta de la inteligencia en la cognición corporizada. Ambas nociones se explican con más detalle en el Cap. 3, Sec. 2.3 de este trabajo.

¹⁷ No por nada la Tesis *Church – Turing* plantea que cualquier algoritmo puede ser entendido como una máquina de Turing (Church, A., (2001) *Logic, Meaning and Computer: Essays in Memory of Alonzo Church* Ed. Anthony Anderson y Micahel Zeleny. Dordrecht: Kluwer Academic Publishers).

La propuesta de Turing se puede ver ahora de manera completa. Obviamente lo que él tenía en mente para el *juego de la imitación* era una máquina con estas naturalezas. Por eso mismo los grandes hallazgos del siglo XX en cuanto a IA y ciencias cognitivas fueron de naturaleza *computacional*.

2.1.2 – Hipótesis de sistema de símbolos físicos

Newell y Simon (N. y S., en adelante) llevaron a cabo una propuesta que sería fundamental para la comprensión del problema de la cognición y la inteligencia durante los siguientes años. Fieles a la metodología de Turing, ellos postularon que las condiciones para que la inteligencia acontezca debían residir en algún grado en la *manipulación simbólica*. Para llegar a esto, N. y S. dan primero una definición de lo que ellos llaman un *sistema de símbolos físicos* (desde ahora SSF):

“Un sistema de símbolos físicos consiste en un set de entidades, llamadas símbolos, los cuales son patrones físicos que pueden ocurrir como componentes de otro tipo de entidad llamada expresión (o estructura simbólica). Así, una estructura simbólica está compuesta de un número de instancias [tokens] de símbolos relacionados de algún modo físico (así como una instancia estando junto a otra)”. (Newell y Simon, 2006, pág. 85 – 86) (Trad. Autor)

Este sistema cuenta además con una serie de procesos que son capaces de manipular las expresiones, tales como crear, modificar, reproducir y destruir símbolos. Finalmente el propósito de un sistema de símbolos físicos es “...producir a través del tiempo una creciente colección de estructuras simbólicas” (Newell y Simon, 1976, pág. 86) (Trad. Autor). El SSF es entonces una máquina que manipula componentes físicos. Estos componentes poseen un significado (son expresiones de algo) en la medida que se conectan entre ellos de una manera física para designar un objeto. El sistema debe *interpretar* la expresión que *designa* un objeto que, a la vez, es otro

símbolo o expresión. Esta interpretación lleva consigo un proceso de significación que es determinado por la misma interpretación y el objeto referido.

Pues apreciarse hasta aquí que la construcción de un SSF representa un método de llevar a cabo tareas de una manera general, es decir, al igual que la máquina de Turing, la hipótesis no provee a la máquina de un propósito específico, necesita ser programada para “algo”, para así determinar los contenidos de las expresiones. Es a partir de este sistema que N. y S. presentan la siguiente hipótesis:

“Un sistema de símbolos físicos tiene los recursos necesarios y suficientes para la acción inteligente general” (Newell y Simon, 1976, pág. 87) (Trad. Autor)

Podemos establecer ahora una diferencia sustantiva entre las hipótesis de Turing y la de N. y S. La hipótesis sostenida por Turing puede ser estipulada de la forma:

(HT) Una máquina que sea capaz de engañar a una persona haciéndose pasar por una, puede ser considerada inteligente.

La hipótesis de Turing tiene claramente un aspecto condicionante: es una conducta ejecutada por la máquina la que la hace inteligente. En cambio, la hipótesis sostenida por N. y S. es metodológica: el SSF posee la capacidad de dar cuenta de una máquina inteligente. Hay que recordar que el mismo Turing hace explícito que la metodología que se use para el desarrollo de la investigación podría ser cualquiera, pero su resultado debía satisfacer la condición del juego de la imitación. Una buena pregunta para hacerse en este respecto sería si N. y S. estarían de acuerdo en decir que el SSF tiene las condiciones necesarias y suficientes para pasar el test de Turing, por lo tanto, satisface esa condición para la *inteligencia*, o, que aunque no llegara a pasar el test de Turing, los recursos del SSF son necesarios y suficientes para la inteligencia de cualquier forma.

Creo que aceptar la segunda interpretación de la hipótesis sería en algún grado, científicamente preocupante. Y no porque no les importe dar cuenta del juego de la imitación, sino porque ese criterio, dar cuenta del juego de la imitación, es de *naturaleza explicativa* y no metodológica. Es decir, hasta el momento, según lo que podemos extraer de SSF, inteligencia sería sólo aquello para lo que la SSF tiene los recursos necesarios y suficientes para dar cuenta, dejando como aspecto relevante de la investigación el método y no el propósito de ésta. La crítica que se le puede hacer a esta manera de interpretar la hipótesis es que, al no establecer un criterio *anterior* a la teoría desde donde podamos establecer el alcance explicativo de la misma, el método es *autosuficiente* explicativamente (sea o no sea este criterio algo así como el test de Turing)¹⁸. Para N. y S., la *necesidad* y la *suficiencia* son propiedades netamente *ingenieriles* del SSF; no debemos pensar que esas propiedades se le conceden mágicamente al sistema de manera *a priori*. N. y S. creen realmente que el sistema da cuenta de manera exitosa, en la experimentación, de lo que llaman *inteligencia general* (a continuación se hablará de esto).

Ahora, N. y S. mencionan a qué se refieren con “acción inteligente general”:

“Por acción inteligente general deseamos indicar el mismo tipo de inteligencia que vemos en la acción humana: que en cualquier situación real, un comportamiento apropiado a los fines del sistema y adaptativo a las demandas del ambiente puede ocurrir, dentro de ciertos límites de velocidad y complejidad” (Newell y Simon, 1976, pág. 87) (Trad. Autor)

Se puede ver que la noción de inteligencia que manejan N. y S. no es para nada muy diferente a la que se intenta defender en esta tesis. Sin embargo, frente a lo expuesto, cabe hacer las siguientes consideraciones a lo citado:

¹⁸ Esto está ligado a la discusión que se da en el capítulo 3 sección 1.1 de este trabajo, acerca de lo que es una *mitología*. Si bien, un investigador es libre de usar las preconcepciones que a él le parezcan necesarias para sostener su teoría, estas preconcepciones deben brindar una explicación satisfactoria del fenómeno en cuestión. Aunque la suficiencia y la necesidad sean concepciones ingenieriles, tienen un peso explicativo que cumplir en la teoría.

- i) Si bien en prácticamente todos los aspectos importantes, la noción de inteligencia de N. y S. se adecúa con la de las teorías alternativas, la consideración de que los recursos del SSF sean suficientes y necesarios para la instanciación de la inteligencia no es una postulado *ex post facto*. Hay varias observaciones que dejarán más que en duda la capacidad de este sistema de cumplir con los requisitos para la inteligencia. Dos de estas observaciones serán abordadas a continuación con los nombres de *problema de la Relevancia* y *problema del Dinamismo*.
- ii) La hipótesis tiene un carácter *cognitivo* en el sentido clásico discutido en el primer capítulo: identifica lo inteligente con el comportamiento de los seres humanos (revisar cita anterior), ya que usa como base que, lo que sea la inteligencia, es llevado a cabo por una *computadora*.
- iii) Si bien la noción expresada hace mención del *ambiente* como punto determinante de lo que es un organismo inteligente, en el testeo experimental la mayoría de (creo que debería decir *todos*) los organismos máquinas modelados por BAIA lidian con *ambientes reducidos*, es decir, son condicionados a ambientes de laboratorio, donde, como dijimos antes, no es posible saber si un organismo es realmente inteligente.

Hechas estas consideraciones me dispongo a revisar ahora dos problemas que han identificado tanto Michael Anderson (Anderson, 2003) como John Tienson (Tienson, 1995) con respecto a la adopción del SSF como método de investigación y creación de IA.

2.1.3 – Problema del Dinamismo¹⁹

En el primer capítulo, se identificaron a la *concordancia* y a la *diversidad* como nociones intuitivamente satisfactorias para individuar organismos inteligentes. Este último criterio, la *diversidad*, tiene un requisito llamado *dinamismo*. Para que un organismo pueda tener la capacidad de comportarse de manera diversa, éste debe ser *dinámico*, lo que significa, en palabras simples, que el organismo debe poder *adaptarse* al ambiente en el que se encuentra, sin necesidad de que el experimentador deba incidir en la máquina para que esto ocurra.

Una máquina basada en los mecanismos teóricos expuestos anteriormente (es decir, símbolos y procesos de manipulación principalmente) necesita necesariamente de un *modelaje*. Un modelo es el programa bajo el cual la máquina es capaz de percibir el mundo (el ambiente), específicamente, el mundo determinado por el modelo. Metafóricamente podríamos decir que la máquina no “vive” de hecho en el mundo real, sino que vive (percibe, delibera y se comporta) en su modelo. Brooks llama a este *marco teórico*, en el que funcionan estos organismos (robots), *marco del sentir/modelar/planear/actuar* (SMPA) (Brooks, 1991a, p. 2). La razón es que las máquinas que funcionan a través de este marco deben sentir (percibir) el ambiente, modelarlo (traducir los datos preceptuales en perceptos los que tienen contenido, es decir, dicen algo acerca del mundo modelado), planear o deliberar su conducta para finalmente llevarla a cabo.

Michael Anderson y John Tienson identifican un problema para los organismos que se ciñen a este tipo de marco (Anderson 2003, Tienson 1995) y que a la vez procuren tener un comportamiento diverso ante el ambiente. Este problema es llamado del *dinamismo*. El ejemplo que da Tienson es muy ilustrativo respecto al problema:

¹⁹ Tienson llama a este problema del *Marco Teórico* [*Framework*]. Si bien se podría decir que hay varias diferencias en la forma en que lo exponen, considero que lo que se quiere dar a entender con ambas exposiciones es equivalente y apuntan a la misma falla del SMPA.

“Considérese el problema de llegar a una tienda específica en el centro comercial. Uno tiene que trazar un trayecto desde la puerta del centro comercial hasta el negocio, usando quizás escalones o una escalera mecánica, evitando los obstáculos estacionados, como plantas, fuentes, bancos... evitando también tropezar con gente... Hacemos esto de manera tan natural que es difícil apreciar la complejidad de la tarea. Pero programar un computador para que lo haga requeriría reglas que tomaran en cuenta una enorme cantidad de factores. El número de computaciones requeridas se expande exponencialmente en relación con el número de factores, conduciendo a una explotación computacional inmanejable.” (Tienson, 1995, pág. 363)

Una máquina que quisiera sortear una tarea como la indicada en el ejemplo es, en teoría, *improgramable* para poder resolver todos los problemas que le pudieran aparecer en el camino. Las razones de por qué una persona puede llevar a cabo esta tarea sin ningún gran problema, es porque las restricciones que se dan en el ambiente son *múltiples* y por lo tanto *débiles*²⁰ para el organismo. Esto quiere decir que el ambiente es de naturaleza impredecible²¹ por lo que cualquier organismo que funcione dentro de él debe tener un *marco suave* que también se puede entender como no tener ningún marco teórico en absoluto. Así, un organismo puede afrontar la variabilidad del ambiente dada sus capacidades estructurales no rígidas. El ejemplo que usa Tienson, luego de la cita anterior, es el del lenguaje. Nosotros si bien, hablamos español, hablamos un dialecto particular de éste. Esto no nos priva de poder entender otros dialectos aunque estos violen las leyes de nuestro lenguaje, “...uno usa las reglas del lenguaje propio para comprenderlo.” (Tienson 1987, pág. 364) Los seres humanos, al igual que todos los organismos biológicos, no nos comportamos bajo reglas específicas de manera que *categorizamos* lo que captamos

²⁰ Restricciones múltiples quiere decir que las leyes que el organismo usa para comprender el mundo están sometidas a constantes relaciones en donde las leyes no se cumplen. Para que un organismo pueda lidiar con estas restricciones debe poseer leyes débiles las cuales, al ser violadas, no desencadenen el colapso del organismo (ver el ejemplo de Tienson mencionado a continuación).

²¹ Entiendo ambiente como el mundo real y no como algún ambiente de laboratorio. En éste último el ambiente sí se torna predecible.

del ambiente. Por el contrario, captamos la *información relevante* del ambiente en la medida que es útil para la conducta. Cuando vamos al centro comercial, ni siquiera somos conscientes de los procesos de “esquivamiento” de obstáculos que llevamos a cabo, al igual que cuando caminamos por una calle llena de gente, postes de luz, intersecciones, etc. Pareciese explicativamente inverosímil postular que cada vez que debemos ejecutar alguna conducta debido a una nueva porción de información obtenida desde el ambiente, lo que hacemos (aún inconscientemente) son computaciones. En cambio, si concediéramos una explicación que postulara que los organismos perciben su ambiente y actúan dada la información que reciben de éste de manera directa, el problema del dinamismo se soluciona.²²

2.1.4 – Problema de la Relevancia

El problema de la relevancia es en rigor, la razón de por qué no es posible solucionar el problema del dinamismo. Un computador no puede captar la información que le resulta *relevante* para la tarea que debe llevar a cabo. La razón es, de nuevo, el número absurdo de computaciones que debería ejecutar para ello:

“Por otro lado, uno no desea reprogramar al sistema en vista de *cada cambio* [en el ambiente], sino, sólo de aquellos que son relevantes, y esos son, los que probablemente puedan afectar la realización de la meta. Así, para un robot pesado que se mueve a través de un cuarto, la localización y dimensiones de los objetos grandes y sólidos es bastante relevante, pero la dirección y velocidad de la corriente de aire que entra por la ventana abierta no lo es...En general, lo que cuente como un hecho relevante que valga la pena ver – es decir, si algo cae o no en la clase general de “obstáculo” – dependerá tanto de las capacidades del agente como de la tarea a realizar. Esto es obviamente suficiente, de forma segura, pero resulta ser notoriamente difícil de

²² Ver cap. 3 sec. 1.2

implementar en un sistema representacional de cualquier tamaño” (Anderson, 2003, Pp. 97 – 98) (Trad. Autor)

La pregunta planteada parece ser ¿cómo podemos programar a una máquina para captar lo relevante, cuando parece que para captar lo relevante uno debe saber de antemano que lo es? La información que pueda estar disponible dentro de un sistema (una memoria) es limitada por una cuestión física. En cambio la información que se encuentra en el ambiente y que resulta relevante para la actividad de los organismos es variable y, por lo mismo, infinita. Hay que mencionar también que si programásemos una máquina para integrar nueva información de manera indiscriminada, al funcionar la máquina bajo un *marco conceptual rígido*, esto podría significar la alteración de toda la estructura de “creencias” de la máquina. Así, la máquina no poseería una visión real del mundo, sino, solamente un modelo del mundo que está en constante cambio. El costo de solucionar de esta manera este problema sería que, otra vez, la máquina no podría saber cuándo es necesario alterar la estructura de creencias y cuándo no, puesto que para esto debería poseer algo así como un marco teórico del marco teórico, lo que nos lleva a una regresión *ad infinitum*.

Tienson (1995, pp. 363 - 366) llama a este problema *del cruzamiento [folding problem]*, y aunque en casi todos los aspectos es equivalente al expuesto por Anderson, Tienson lo ve desde el punto de vista de la imposibilidad de una máquina de cruzar guiones [*scripts*] de comportamientos. Un guión, sería la manera en que una máquina podría contextualizar ciertos tipos de información para que el procedimiento heurístico no sea tan quimérico. Así, una máquina que se encuentra en un cumpleaños, puede identificar sólo partes de la información y así inferir el guión que debe utilizar. El problema de la relevancia se da aquí de la misma manera que en el caso anterior: una máquina no sería capaz de *cruzar* dos o más guiones en caso de ser necesario. ¿Qué tal si hubiera un incendio en el cumpleaños? Bueno, el robot debería perfectamente poder cruzar los guiones de “incendio” y “cumpleaños”. Pero ¿y si el incendio es en una casa de dos pisos?, ¿Podrá cruzar los guiones “incendio”,

“cumpleaños” y “casa de dos pisos”?, ¿Y si además hay gente discapacitada? Parece sensato pensar que el cruzamiento tendrá en algún momento un límite, ya sea porque la máquina será incapaz de cruzar todos los guiones necesarios, o simplemente porque esos guiones no estarán disponibles (pensar que se pueden almacenar en una memoria todos los guiones de conducta posible, es igual de ilusorio que pensar que todas las variables del ambiente pueden ser integradas).

Desde este punto de vista es altamente dificultoso lograr dar una solución de este problema desde el marco SMPA.

2.1.5 - Evaluación

A continuación realizaré dos observaciones acerca de los contenidos de la primera parte de este capítulo.

- 1) El *Test de Turing* parece no ser, en un sentido general, una mala meta para la investigación de la IA y por lo tanto, para cualquier programa de investigación que exista en esta disciplina. Lo que sí es por lo menos discutible es el enfoque con el que deba ser mirado este *juego de la imitación*. Si bien, que una máquina logre hacerse pasar por una persona no es un indicador incorrecto de *inteligencia*, sigue teniendo el carácter humanista discutido en el primer capítulo. Lo que se sostendrá en el capítulo número tres será que este juego de la imitación puede ser un mejor test para la inteligencia si consideramos que el engaño se debe producir en el *mundo real* y no en ambientes controlados. Por otra parte el juego no sólo debe incluir la imitación de personas, sino el de cualquier organismo al que podamos adscribir inteligencia según los objetivos planteados en el capítulo uno. Así, no deberíamos decir que el organismo (máquina, robot) se comporta como imitando a algún animal específico, sino, que se comporta como imitando a un *organismo biológico*.
- 2) La hipótesis de sistema de símbolos físicos es la esencia de toda máquina computadora. Por lo mismo, al necesitar de modelación, se le pueden atribuir

los dos problemas señalados anteriormente: el de la *dinamicidad* y el de la *relevancia*. Como se pudo ver, ambos problemas resultan sumamente perjudiciales en el objetivo de crear un organismo que pueda ser llamado propiamente inteligente. De esta manera, es posible pensar que cualquier tipo de solución a estos dos problemas puede ser mejor abordada desde las propuestas de los *programas alternativos* a BAIA. De lo contrario, lo único que podemos esperar es una postergación del problema sobre soluciones parciales. Creo que, ante lo que muestra la evidencia, ambos problemas no son ni esencialmente técnicos ni metodológicos, sino de *principios*: la razón de por qué la IA clásica no ha podido dar cuenta de estos problemas es porque la metodología con que trabaja se basa en el fundamento de que los procesos computacionales son indispensables para la acción inteligente.

Ahora en la siguiente sección se abordará el planteamiento del *Conexionismo* como programa de investigación y como posible solución a los problemas del dinamismo y de la relevancia.

2.2 - Conexionismo

El conexionismo es otro de los intentos de dar una explicación a la cognición, y, por consiguiente, una solución al problema de la inteligencia. Es quizás la postura más desafortunada respecto al lugar que ocupa dentro de las ciencias cognitivas, puesto que las críticas que se le hacen vienen tanto del lado de los cognitivistas clásicos como del de los defensores de las teorías alternativas. Estas críticas sin embargo no son para nada vacías, y tienen razón de ser. A continuación, se hará una breve descripción de lo que es un sistema conexionista, caracterizando sus partes fundamentales y las razones que los partidarios de este programa tienen para considerarlo diferente de la postura clásica. Para esto me basaré fundamentalmente en las descripciones de Tienson (Tienson, 1987), Smolensky (Smolensky, 1988) y Rumelhart (Rumelhart, 1989) de los sistemas conexionistas. Finalmente, se verá si el paradigma conexionista es capaz de solucionar los problemas del dinamismo y la relevancia.

2.2.1 – Sistema Conexionista

Un sistema conexionista es una red interconectada de *nodos*. Estos nodos tienen dos estados de activación (encendido, apagado) y para poder ser activados deben ser afectados por otro nodo. Cada nodo posee un umbral de activación²³, es decir, un número según el cual el nodo se encenderá o apagará (si el nodo recibe una señal de menor intensidad al número del umbral, entonces no se activará). La activación del nodo depende de ciertas propiedades de la señal tales como su intensidad y de la fuerza de conexión (*peso*) que haya entre el nodo remitente y el receptor. Esta red intenta emular la forma de las *redes neurales* a manera de asemejarse a los mecanismos cerebrales. Una de las características de esta red es que los procesamientos que se llevan a cabo dentro de ella ocurren en paralelo y no de manera lineal. En la arquitectura clásica, la estructura interna de las máquinas es la de un computador Von Neumann: una línea algorítmica *serial*. Allí, el *input* perceptual pasa a las unidades de procesamiento internas que luego dan lugar al *output* conductual. El procesamiento en paralelo resulta una característica necesaria para el sistema conexionista, dado que, por la naturaleza de sus nodos, la información que el sistema alberga se encuentra distribuida de manera aleatoria, dependiendo de los *pesos* de las conexiones entre los nodos. Los *pesos* determinan el tipo de conexión que se da entre los nodos. En la medida que un sistema conexionista es sometido a pruebas a fin de que pueda aprender a ejecutar cierta tarea, la manera en que el sistema aprende es “engrosando” las conexiones de los nodos que dan las respuestas correctas a los problemas planteados. Así, si dos nodos poseen una conexión muy fuerte es más probable que cuando uno esté activado el otro también lo esté. Tienson llama a este tipo de conexión *excitatoria*. Así, un peso negativo entre dos nodos hará que cuando uno esté activado, el otro nunca lo esté. Esta conexión es llamada *inhibitoria* (Tienson, 1987). De esta manera, se podría decir que en los pesos se

²³ Puede no poseerlo. Pudiera ser que el nodo se activara o desactivara simplemente al ser afectado por otro.

encuentra el *programa* del sistema²⁴. Existen ciertos nodos que captan el *input* externo, para después pasar la información a los nodos que les corresponda según la distribución de los pesos (estos nodos reciben el nombre de *unidades escondidas*). Finalmente, una última capa de nodos genera el *output* conductual.

Para que el sistema pueda de hecho aprender debe existir un *instructor* que corrija al sistema cada vez que este ejecute un comportamiento errado. Según Rumelhart (Rumelhart, 1989) existen tres maneras en que el sistema puede lograr aprender: (1) Desarrollando nuevas conexiones, (2) perdiendo conexiones existentes y (3) modificando los pesos de las conexiones que ya existen. Rumelhart cree que si bien, muy poco trabajo se ha hecho en (1) y (2) hay un aspecto en que (3) es equivalente a (1) y (2):

“Donde quiera que cambiemos el peso de una conexión de cero a algún valor positivo o negativo, tiene el mismo efecto de crear una conexión nueva. Donde quiera que cambiemos el peso de una conexión a cero, aquello tiene el mismo efecto que perder una conexión.” (Rumelhart, 1989, p. 213) (Trad. Autor)

El proceso de volver a distribuir los pesos de las conexiones luego del ensayo y el error, se llama *retropropagación*, y consiste básicamente en la *recalibración* de los pesos según el grado de éxito obtenido en la tarea. Si la máquina no logró llevar a cabo la tarea propuesta, la recalibración funciona debilitando la conexión que se dio entre los nodos que procesaron el input. Así, en el futuro, la máquina privilegiará cualquier otra conexión antes que esa.

²⁴ Un programa o *software* en una máquina que funciona al estilo Von Neumann, tiene la función de establecer qué conexiones de las que se dan en el sistema son las relevantes para los procesos a ejecutar, o, de qué manera lo son. Así, el programa establece un límite, un *mundo de conexiones significativas*. En la máquina conexionista esto no se encuentra explícito, ya que la máquina no es programada para ninguna tarea particular en un comienzo. El adiestramiento de la máquina moldea los pesos de las conexiones entre los nodos, haciendo a unas más fuertes y a otras más débiles. Así, cuando un input perceptual es procesado de manera sub - simbólica, pasa a través de los pesos ya establecidos como lo haría el mismo input a través de las conexiones previamente establecidas en una máquina Von Neumann clásica.

2.2.2 – Nivel conceptual v/s nivel sub-conceptual

La manera en que los nodos actúan en la explicación psicológica conexionista es a la manera de *sub-símbolos*. Los nodos son los componentes de los conceptos. De esta manera la explicación se reduce un nivel, y esta podría así (según Smolensky 1988) tener una mejor descripción de los procesamientos llamados *intuitivos*. Entonces, cuando uno estuviera procesando o entendiendo un concepto como “el perro dálmata de mi tía Beatriz” los nodos actuarían como constituyentes de esa descripción. Éstos, sin embargo, no son *constituyentes semánticos* de la oración mental. Los nodos actúan en un nivel diferente al del “lenguaje del pensamiento”, están sometidos a las reglas de la matemática:

“Las entidades que son representadas típicamente por símbolos en el paradigma simbólico, son típicamente representadas por sub-símbolos en el paradigma sub-simbólico. Junto con esta distinción semántica viene también una distinción sintáctica. Los sub-símbolos no son operados por manipulación simbólica: ellos participan en la computación numérica no simbólica.” (Smolensky, 1988, p. 774)

Por este motivo Smolensky llama al nivel simbólico *conceptual*, y al sub-simbólico *sub-conceptual*. La razón de por qué se busca reducir la explicación a este nivel sub-conceptual es porque los procesos intuitivos, de percepción y todos aquellos que no parecieran propiamente “cognitivos”, no poseen una explicación satisfactoria desde el paradigma simbólico. Recordemos el ejemplo de Tienson en referencia al problema del dinamismo. Cuando la máquina va al centro comercial debe evitar un montón de obstáculos de una manera no cognitiva (de una manera más bien, reactiva). En este caso, dicho comportamiento tendría una mejor explicación si la reducimos a un nivel sub-simbólico: la máquina percibe el ambiente y en lugar de realizar un proceso heurístico para buscar información y un proceso de manipulación simbólica para deliberar y ejecutar su comportamiento, lo que hace es captar el *input* y dejar que la señal viaje a través de las conexiones ya establecidas (los pesos ya están distribuidos),

sin necesidad de recurrir a la manipulación simbólica. En este caso, el proceso sub-simbólico sería mucho más rápido y parecido al proceso humano al generar una conducta similar en un tiempo también similar.

2.2.3 - Soluciones al problema del dinamismo y al problema de la relevancia

En el ejemplo anterior pudimos ver que, en teoría, el sistema conexionista tendría un mejor desempeño frente al problema del dinamismo que el que tendría el cognitivismo clásico y ortodoxo. Sin embargo, no es sólo una cuestión de rapidez el poder dar cuenta de este problema. La multiplicidad de tareas simultáneas es también algo que está presente en nuestra concepción de dinamismo²⁵.

El efecto de *retro-propagación* se ha caracterizado por ser un proceso sumamente lento. Una máquina necesita repetir una tarea (sencilla, por lo demás) demasiadas veces para poder dar con la *conexión correcta*. Los programas que sirven de ejemplo en las descripciones de los sistemas conexionistas son siempre de naturaleza simple: programas para reconocer rostros, para identificar miembros de un árbol genealógico, etc. y aún así los procesos de aprendizaje son lentísimos (Tienson 1995, Somlensky 1988). Imaginemos que necesitásemos una máquina capaz de identificar razas de perros, marcas de auto, diferenciar colores primarios de los demás y sumar números naturales. Al no tener los sub-símbolos contenido semántico del tipo de las descripciones simbólicas (pero teniendo, de todos modos, la intención de dar cuenta de ese tipo de explicaciones, es decir, de la aprehensión de *conceptos*) los nodos ocupados en todos esos procesos son los mismos. Intuitivamente la cantidad de conexiones que debería haber sería demasiada, y por lo tanto, el proceso de aprendizaje sería interminable.

La cuestión fundamental que plantea el problema del dinamismo es que una máquina que funcione bajo un marco teórico del tipo SMPA o i) caerá en una explosión

²⁵ La importancia de esta multiplicidad se encuentra detallada en el cap. 3, sec. 2.5 y siguientes.

computacional ii) o no podrá dar cuenta de la variabilidad del mundo. Frente a (i), el conexionismo, al parecer, no tiene problemas, pero lo que hace realmente la propuesta es intercambiar este problema por el de la “explosión” de conexiones explicado más arriba. El segundo punto (ii) es también una incógnita: programar una máquina conexionista para tareas múltiples y simultáneas es una tarea más que pendiente, y, a la luz de lo que acabamos de decir acerca de (i), creo que la respuesta a este problema sigue siendo negativa. Es lamentable que dada la juventud del programa conexionista en el año de la publicación de su artículo, Tienson mismo (Tienson 1995, original de 1986) que menciona el problema del *marco teórico*, no pudiera dar una explicación con más detalle de porqué el conexionismo es una propuesta que lo resuelve. El problema del *cruzamiento* es mejor abordable desde el conexionismo, puesto que los procesos heurísticos que se realizan en los sistemas simbólicos no existen en los sistemas conexionistas. El sistema conexionista puede dar con lo relevante si es que es adiestrado lo suficiente en ello, sin necesidad de almacenar memoria conceptual que le permita identificar contextos diferentes.

2.2.4 -Evaluación

Ya hemos visto en líneas generales de que se trata el conexionismo y cuáles son sus metas como programa. Las posturas conexionistas en general, no pretenden dar explicaciones que reemplacen al paradigma clásico de suyo, sino que pretenden reemplazarlo ahí sólo donde es necesario:

“La relación entre los modelos sub-simbólico y simbólico es como la que existe entre la mecánica cuántica y la clásica. Los modelos sub –simbólicos describen de manera fiel la micro – estructura de la cognición allí donde los modelos simbólicos proveen una descripción aproximada de la macro-estructura. Una tarea importante la teoría sub-simbólica es delinear las situaciones y los respectos en los cuáles la aproximación simbólica es válida y explicar por qué...” (Smolensky, 1988, p. 793) (Trad. Autor)

Hay evidentemente un lugar para la explicación cognitivista y otro para la explicación conexionista. Sin embargo, que la relación que existe entre ambos paradigmas es como la existente entre la mecánica cuántica y la clásica no se sigue: las explicaciones de las redes conexionistas al nivel cognitivo parecen irrelevantes si no se entienden como la antesala de la explicación simbólica, la mecánica cuántica, en cambio, tiene valor explicativo propio. Brooks hace también un alcance con respecto a esto mismo:

“...ellos [los conexionistas] están buscando revelaciones desde el entendimiento de cómo conectarlos [los nodos] correctamente. ...los conexionistas parecen estar buscando representaciones distribuidas explícitas que espontáneamente aparezcan de sus redes.” (Brooks, 1991b, p. 9) (Trad. Autor)

A pesar de que una de las ventajas del sistema conexionista es que la máquina no necesita de todos los sub-símbolos para poder inferir el/los restante(s) (Tienson, 1995), el hecho de que deba inferir algo(o quizás el simple hecho de llamar a los componentes “sub-símbolos”) para poder dar cuenta de la instanciación correcta, muestra que el conexionismo intenta explicar las fallas del cognitivismo clásico pero no sin él. Es más que posible entender el conexionismo como una forma de instanciar el paradigma clásico, puesto que no se preocupa de alejar lo suficiente la relevancia de la *explicación simbólica* para dar paso a una explicación matemática, más bien intenta dar con los “símbolos” desde un nivel inferior:

“Donde el cognitivismo ortodoxo clásico asigna contenido simbólico al tipo de entidades físicas que almacenan las arquitecturas Von Neumann, el cognitivismo conexionista asigna contenido sub-simbólico al tipo de entidades físicas que están distribuidas y superpuestas en la matriz de pesos de la red” (Calvo & Gomila, 2008, p. 4) (Trad. Autor)

Finalmente, y dada la discusión propuesta en el capítulo uno de esta tesis, haré otras dos consideraciones:

- 1) La concepción acerca de lo que es el *ambiente* y, el rol que cumple dentro de la teoría conexionista, están basados en un punto de vista incorrecto si lo que se busca es un paradigma que supere los problemas del cognitivismo ortodoxo. Por ejemplo, Rumelhart caracteriza el ambiente como: "...una función estocástica de tiempo variante sobre un espacio de posibles patrones de input." (Rumelhart, 1999, p. 215) (Trad. Autor). La función es acerca de "posibles patrones de input", lo que implica que sea una función estadística que no ocupa al mundo real como modelo, sino a la función, a los "posibles patrones de input". Esto ya implica una categorización impropia para superar el marco SMPA. Con respecto a esto mismo, al no haber una máquina conexionista que funcione en el mundo real, los posibles patrones de input son finalmente los que los ingenieros puedan imaginar, pero no los que realmente hay. Este último punto es muy importante para la desestimación de la teoría conexionista por parte de las teorías alternativas de la cognición. Una máquina conexionista posee como ambiente una función dentro de la cual se incluyen los posibles patrones de input que puede captar. No es posible llevar a cabo una máquina con esa metodología de percepción en el mundo real, puesto que el ambiente es cambiante, y no hay función que permita coleccionar todos los posibles patrones de input, a menos, que entendamos esta colección de la misma manera que entendemos la percepción a través del cognitivismo clásico ortodoxo, y esto quiere decir, percibir con *conceptos*.
- 2) Una posición que quiera dar una solución real a los problemas del dinamismo y la relevancia para poder alcanzar la concepción de inteligencia que persigue la ciencia cognitiva y sobre todo la IA (vista en el primer capítulo) debe poder dar cuenta de *comportamientos múltiples y paralelos*, como los que poseen las personas y los animales. De poco nos sirve que el conexionismo logre construir una máquina capaz de reconocer rostros si: (i) difícilmente puede ser programada para otras tareas, (ii) no cuenta con el mundo real como modelo y (iii) no hay una versión de una máquina (robot) situada en el mundo real que

pueda ser una base para la investigación de esta postura en lo que respecta a los problemas de la cognición y la inteligencia.

En el siguiente capítulo se abordará la parte central de esta tesis que trata acerca de cómo la cognición corporizada ofrece una concepción de la inteligencia que: 1) es intuitivamente satisfactoria para la explicación de la conducta animal, humana y de la IA, 2) resuelve teóricamente los problemas del dinamismo y la relevancia y 3) ofrece una mejor comprensión del fenómeno de la cognición. Específicamente, la perspectiva de la cognición corporizada desarrollada en ese capítulo, es la propuesta por el ingeniero en robótica, Rodney Brooks. Antes, a modo de introducción, se expondrán algunos contenidos que pueden ser considerados como antecedentes históricos para la propuesta de la cognición corporizada. Así mismo, se hará también un bosquejo general de lo que buscan las teorías alternativas al paradigma clásico.

Capítulo 3 - Post-cognitivismo: Cognición Corporizada

En este capítulo se exhibirá la propuesta central de este trabajo. Ya se ha expuesto en los capítulos anteriores de lo que trata, a grandes rasgos, la noción de inteligencia que se muestra la teoría cognitiva ortodoxa y en la teoría cognitiva conexionista. En resumen, se vio que cualquier concepción de lo que sea *inteligencia* que podamos recoger desde el *cognitivismo* debe estar basada en lo que Newell y Simon llaman *hipótesis de sistema de símbolos físicos*, ya sea de manera explícita como en el cognitivismo clásico, o de manera implícita como en el conexionismo. También vimos que aceptar esta hipótesis nos lleva a una concepción de inteligencia que es determinada por los mecanismos internos de los organismos. Este criterio para dar con la inteligencia es, como ya vimos, no tan satisfactorio como sus defensores podrían llegar a creer, y en la medida de solucionar los problemas revelados en el capítulo anterior, es también contrario al planteamiento del primer capítulo, de acuerdo al cual que los criterios por los cuales la inteligencia debe ser reconocida han de ser de orden epistémico, inter-subjetivo y no internistas. Por esto último es que se identifican como criterios relevantes para la identificación de comportamientos inteligentes la *diversidad* y la *concordancia*.

3.1 – Consideraciones Preliminares

Antes de entrar a una descripción de la robótica de Rodney Brooks y lo que su ingeniería plantea, se expondrán algunas consideraciones que son muy importantes para entender cómo es que los partidarios de las teorías alternativas de la explicación de la conducta han llevado a cabo sus planteamientos y dónde reside su fundamento filosófico.

3.1.1 – La mitología cognitivista

Es útil para el entendimiento de la siguiente sección de esta tesis preguntarse cuáles son las intenciones con las que abordamos las diferentes propuestas para una

explicación de la inteligencia. Quiero decir que, es menester preguntarse en qué lugar se pone la crítica que se ha hecho hasta ahora acerca del paradigma clásico. Siguiendo el método wittgensteineano de Rowlands (Rowlands, 1999, p. 8) diré que hay, al menos, dos maneras de hacerle una crítica a un paradigma de explicación. Una es desde dentro del paradigma mismo, mostrando que es incoherente internamente o que las creencias que lo fundan son falsas. Algo de esto se vio ya en el capítulo anterior, en el que se consideró cómo se dan los problemas que se caracterizaron como problemas de principio. La otra manera es, simplemente, hacer notar que los conceptos pre - teóricos en los que se basa un paradigma de explicación no son necesarios para toda clase de explicación de un mismo fenómeno. Rowlands usa la noción wittgensteineana de *mitología*, para identificar a aquellas nociones que se ubican antes del paradigma y que sirven para sentar las bases de “cómo deberían ser las cosas”, es decir cómo deberíamos entender “mente”, “cognición”, “inteligencia”, etc. Una cosa es que la explicación, ya basada en estas nociones pre - teóricas, sea inconsistente o, de suyo, falsa. Otra es que, antes de la explicación, simplemente nos preguntemos por qué tenemos que aceptar la “mitología” que se nos ha entregado como marco de investigación. Así, de la segunda manera, sin tener que refutar nada a nadie, se puede comenzar a esgrimir un nuevo paradigma de explicación, sólo diciendo al respecto que podemos aceptar lo que nos dice el cognitivismo con respecto a los mecanismos y las naturalezas de lo mental, o no. De acuerdo con eso, optaré a continuación por la segunda alternativa, y presentaré un nuevo marco (*mitología*) que, dada la manera de entender los problemas y los objetivos de la ciencia cognitiva y la IA que se ha defendido hasta ahora, me parece más convincente y útil para esos propósitos.

Ya se ha hablado de los reparos que pueden hacerse al paradigma clásico. Es por eso que también me gustaría hacer el siguiente comentario: No hay paradigma explicativo en la actualidad que trate los temas aludidos, al que no puedan hacerse reparos desde el otro lado de la cerca. En lo que respecta a las ciencias cognitivas, creo que el paradigma que debería predominar debería ser aquel que se ajuste más a los principios de la disciplina. Como no hay un acuerdo común en eso, por lo menos

ahora, la discusión seguirá por algún tiempo. En la IA la cosa no es muy diferente, pero considero que el nombre de la disciplina alberga un principio mucho más explícito que en las ciencias cognitivas: la IA debe poder lograr la creación de un robot/máquina inteligente, entendiendo inteligencia como aquello que tienen, al menos, los humanos. Concedido esto, el paradigma que se deba priorizar en IA debiera hacer aquel que en la práctica nos entregue más réditos al respecto. Hasta el momento, no creo ver la balanza muy inclinada para ningún lado si es que dejamos la historia fuera de ella. Entendiendo lo anterior, se debería leer la siguiente última parte de esta tesis no como la exposición de un marco de investigación que intenta refutar al marco anterior, sino, como la exposición de un marco de investigación que, dado que la exposición del paradigma clásico mostró que éste puede no ser de alguna manera completamente satisfactorio, posee la coherencia y bases suficientes para asentarse como una buena alternativa para la explicación de los fenómenos de la conducta y la inteligencia.

A continuación, antes de describir la teoría de la inteligencia postulada por Rodney Brooks, se hará una introducción en la cual se intentará rescatar ciertas nociones de filósofos anteriores a los científicos cognitivos actuales. Estas nociones deben entenderse, no como una manera de decir que lo postulado por estos filósofos es de alguna manera correlativo o equivalente a lo dicho por los cognitivistas alternativos actuales, sino, como la muestra de que la intención de dudar de ciertas concepciones pre - teóricas en las que se basan las hipótesis del paradigma clásico en ciencias cognitivas, ya se encontraba anteriormente en los pensamientos de algunos investigadores, y que, estos pensamientos, contenían ideas más que rescatables.

3.1.2 - Percepción Directa y Ambiente

Las teorías de la cognición que se basan en, lo que podríamos llamar, paradigma alternativo en ciencias cognitivas, utilizan varios argumentos que se encontraban de alguna manera en las creencias de filósofos, biólogos y sicólogos de principios del siglo XX. Dos de los investigadores más citados por los científicos alternativos son el

sicólogo James J. Gibson y el biólogo Jakob Von Uexküll. El primero, es el autor de una teoría de la percepción que implica que los procesos que generalmente se le adscriben a la mente con respecto a la interpretación del estímulo perceptual proximal, no eran de hecho necesarios para la explicación de la percepción. Esta visión es sólo necesaria si pensamos que el ambiente no posee la información suficiente para dar cuenta de la cognición²⁶ y por ende, de la acción.

Entonces ¿qué es lo que ocurre? El estímulo captado por el organismo posee de hecho la información suficiente para dar cuenta de la conducta del mismo. Esta manera de entender la percepción implica que la cognición o (i) es un proceso que involucra la apreciación de la acción de un organismo por parte de un testigo, o, como diría Brooks:

“... [el nuevo enfoque] postula que los sub - sistemas de percepción y acción hacen todo el trabajo y que es sólo un observador externo el que tiene que ver con la cognición, por el hecho de atribuir habilidades cognitivas al sistema que funciona bien en el mundo pero no posee ningún lugar específico donde la cognición pueda estar llevándose a cabo” .
(Brooks, 1999, p. 8) (Trad. Autor)

O también (ii) nos podría llevar a aceptar que la cognición de hecho no ocurre de ninguna manera. Eso quizás es suficiente para optar por (i) o por desechar cualquier manera de entender la cognición, con el fin de hacer nuestro marco teórico más simple. De esta manera, la percepción directa es un antecedente obligado para las hipótesis actuales en ciencias cognitivas: percibir la información de manera directa nos hace desestimar los mecanismos internos de los organismos como recursos necesarios para llevar a cabo la acción inteligente.

²⁶ Este es el clásico argumento de Noam Chomsky de la pobreza del estímulo (APE). Las teorías de la percepción del paradigma clásico utilizan este argumento para la postulación de mecanismos internos que medien la percepción y la acción.

Tanto Gibson como Uexküll tienen una noción de ambiente que es común para los investigadores de las teorías alternativas. Uexküll entiende el ambiente (*unwelt*) como la conjunción de otros dos ambientes: el *mundo perceptivo* y el *mundo efectivo* (Uexküll, 2010, p. 42). El primero, es aquel ambiente donde el agente percibe la información que se encuentra en el mundo; el segundo, el ambiente en el cuál se comporta. Estos mundos, según Uexküll, son herramientas explicativas suficientes para dar cuenta de los comportamientos de los animales. De manera similar, la teoría de Gibson postula una relación entre lo que es *percepción* y *acción* (Gibson, 1986). En el *arreglo ambiental*²⁷, se encuentran las propiedades de aquel ambiente, las cuáles acarrearán información acerca de cómo comportarse. Las propiedades que se exhiben en el ambiente y que portan contenido acerca de la conducta son llamadas por Gibson *affordances*. Los *affordances* son *oportunidades de acción* que se determinan a partir del organismo y el ambiente, es decir, son *propiedades relacionales*. De esta manera, como se mencionaba en el primer capítulo de este trabajo, lo que sea que obtengamos del ambiente depende de nuestras capacidades y naturaleza en cuánto organismos; lo que el ambiente nos ofrece es lo que nosotros somos capaces de hacer.

3.1.3 – Operadores de máquina

Uexküll usa de ejemplo el comportamiento de una garrapata para hacer una comparación entre la explicación que pueden dar un *fisiólogo* y un *biólogo* de la conducta del ácaro (Uexküll, 1986, pp. 51 - 52). Una garrapata actúa percibiendo *índices de calor*, los cuales le permiten captar la presencia de organismos de sangre caliente sobre los que puede alojarse y así extraer su sangre. En resumen, una garrapata percibe calor, trepa o se deja caer sobre su objetivo, extrae sangre de éste hasta que no puede almacenar más y repite este comportamiento hasta su muerte. Un fisiólogo (en la explicación de Uexküll) entendería el comportamiento de la garrapata

²⁷ Arreglo ambiental es un concepto introducido por Gibson (Gibson, 1986, p. 80) en su teoría acerca de la percepción visual. Hay que tener siempre en cuenta que en la teoría de la percepción de Gibson, el ambiente es *estructurado*, esto significa que las relaciones que se dan entre, en este caso, el observador y el ambiente están *normadas* por leyes naturales. El arreglo ambiental es aquella serie de estructuras que toman parte en la relación existente entre el sujeto y el ambiente.

como el comportamiento de una máquina, por lo que el comportamiento de las partes de la garrapata es el comportamiento de “pequeñas máquinas” que componen a la grande. Un biólogo en cambio, entendería el comportamiento de las partes de la garrapata (perceptores, actuadores, etc.) como el comportamiento de *operadores de máquina* ya que, lo que ocurre en realidad es una transferencia de estímulo y no de movimiento. Para que haya un comportamiento a través de un estímulo, se necesita de un sujeto (que se entiende como la conjunción de estos múltiples operadores de máquina) que interprete (no en un sentido de interpretación de símbolos) el estímulo perceptual. Uexküll dice que los músculos, por ejemplo, ante cualquier estímulo externo, reaccionan contrayéndose. Éstos toman el signo del estímulo y según sus propiedades físicas lo interpretan como una contracción del músculo (percepción – acción). El escarabajo funciona de la misma manera. Las máquinas, por el contrario, funcionan de acuerdo a su rol establecido. Un botón, por ejemplo, reaccionara de manera adecuada sólo si es presionado. Pero si le echamos ácido, o le prendemos fuego no reaccionará de ninguna manera.

La intención de Uexküll a través de este ejemplo es poder establecer una clara distinción de naturalezas entre el comportamiento de una máquina (mecánica) y el de un ser vivo. Las partes de una máquina no tienen *reacción* realmente, sólo poseen una función específica: son las partes que componen un organismo las que se pueden llamar propiamente *reactivas*. La *reacción* es un principio vital en la construcción de máquinas si nuestro objetivo es que se parezcan en algo a nosotros. Una manera de ver la discusión actual en IA es poniéndola en los términos de Uexküll. La IA clásica se especializa en la construcción de máquinas cuyos componentes son otras pequeñas máquinas de naturaleza mecánica: sólo un tipo de estímulo tiene valor perceptual para la máquina. En cambio, lo que intenta hacer la cognición corporizada (y como se verá más adelante, la robótica de Brooks) es incorporar la *reactividad* en cada componente de la máquina que genere un comportamiento. El “operador de máquina” debe estar *situado* en el mundo de tal manera que cualquier tipo de información que sea captada desde el ambiente pueda ser traducida a un comportamiento *reactivo*.

3.1.4 - Estableciendo bases para una inteligencia corporizada: Concordancia y Diversidad

Lo recién expuesto tuvo como tarea sentar las bases para la discusión de un nuevo paradigma de explicación, o como habíamos mencionado, una nueva *mitología*.

La hipótesis de la percepción directa es una base importante para las propuestas del paradigma alternativo en ciencias cognitivas (cognición corporizada, sistemas dinámicos, enacción, y otras).

Ahora, en el objetivo de estipular una teoría acerca de la inteligencia, se seguirá la línea investigativa propuesta en el primer capítulo. Allí, nos referimos a la importancia de:

- 1) Entender *inteligencia* primeramente como aquello que distingue a los organismos biológicos.
- 2) Que los criterios utilizados para distinguir la inteligencia no deben ser internistas (no deben ser acerca de los mecanismos internos)
- 3) Una teoría de la inteligencia debería abarcar tanto a animales como a personas y máquinas.

Para poder llegar a una primera aceptación intuitiva de (1) comenzaré argumentando lo siguiente. Una teoría que adscriba inteligencia sólo a los humanos (o que adscriba inteligencia a los humanos como esencialmente diferente a la de los animales) debe explicar por qué la inteligencia acontece sólo en los seres humanos. La propuesta natural es pensar que las personas poseen algo así como *cognición* mientras que los animales no. Esto se hace generalmente, argumentando que los seres humanos poseen un lenguaje gramatical, además, de un comportamiento determinado por creencias y deseos (estados intencionales). Todo esto al parecer involucra algo así como la *cognición*, pero por qué lo hace es algo que aún no puede tomarse como evidente. La pregunta ahora es cómo nos hacemos cargo de la distinción animal – hombre. Este punto fue tocado en el primer capítulo cuando se hablaba acerca de la deshumanización de la inteligencia. Siguiendo el espíritu de esta segunda parte de

esta tesis, diré que los criterios expuestos como *concordancia* y *diversidad* para la identificación de la inteligencia son una manera de adoptar un marco de investigación diferente tanto porque no estamos obligados a seguir ligados a la antigua manera de comprender a la inteligencia, como porque en el marco antiguo se ilustraron problemas que perjudicaban los objetivos de la IA y las ciencias cognitivas. Es por eso que no deberíamos hacernos cargo de la distinción hombre – animal, porque buscamos una definición (explicación, etc.) de inteligencia que la supere. Pfeifer y Bongard dicen al respecto:

“De una manera u otra, una teoría debería capturar nuestro entendimiento del campo de una manera compacta, así las intuiciones pueden ser aplicadas a muchos problemas diferentes en el área y pueden ser ampliamente comunicados. Queremos que nuestra teoría no caracterice sólo a hormigas y ratas, sino también a humanos, robots y quizás otros tipos de artefactos como teléfonos móviles, autos inteligentes y poleras cableadas [*wired T-shirts*]” (Pfeifer y Bongard, 2006, p. 84) (Trad. Autor)

Cuando decimos que lo inteligente se caracteriza por ser concordante y diverso, estamos diciendo que lo que pueda instanciar esas características es irrelevante para la explicación de la inteligencia. Si manifiesta una conducta que está en *concordancia* con las reglas sociales y las leyes físicas y es además *diverso* en sus maneras de ejecución, entonces es acreedor de lo que queremos decir por inteligencia.

La parte central de esta tesis es la defensa de la propuesta de Rodney Brooks como un método válido tanto para la explicación de la inteligencia como para la creación de máquinas inteligentes. A continuación se verá que sus criterios acerca de cómo y de qué manera se desenvuelven los organismos inteligentes satisfacen con los criterios de concordancia y diversidad. Para lograr aquella apreciación se desarrollará tanto la parte teórica de su investigación como la parte técnica de los principios en los que basa su robótica. Esto último dado que él mismo dice que su trabajo se “desprende de consideraciones ingenieriles” (Brooks, 1991b, p. 10) (Trad. Autor)

3.2 - La robótica corporizada de Rodney Brooks

Creo que se han ya introducido de manera más que general tanto lo que es la teoría de la inteligencia y de la cognición del paradigma clásico en ciencias cognitivas e IA como los principios bajo los cuales los científicos cognitivos alternativos intentan fundar un nuevo paradigma para la explicación de estos mismos fenómenos. Es por eso que ahora pasaré a detallar tanto la propuesta filosófico – conceptual, como la propuesta técnica del ingeniero en robótica Rodney Brooks para un mejor entendimiento de lo que es la inteligencia así como para la obtención una mejor metodología en la construcción de máquinas “inteligentes”.

3.2.1 – Una conclusión inesperada y una hipótesis radical

Uno de los puntos de partida más importantes en el argumento de Brooks respecto a investigación de la inteligencia se ve reflejado en la siguiente cita:

“... creemos que la inteligencia de nivel humano es demasiado compleja y poco entendida como para, por el momento, ser descompuesta en las sub – piezas correctas, y que, incluso si conociéramos las sub – piezas aún no conoceríamos las interfases correctas entre ellas. Es más, nunca entenderemos como descomponer la inteligencia de nivel humano sin antes tener mucha práctica con inteligencia de nivel más simple.” (Brooks, 1991b, p.1) (Trad. Autor)

A través de lo expuesto en esta cita podemos desprender los siguientes principios de investigación:

(1) Buscar descomponer la inteligencia en los mecanismos que la llevan a cabo no es la metodología más adecuada para llevar a cabo la investigación del fenómeno en cuestión (como lo hace la IA clásica).

(2) Es sensato creer que la práctica con niveles más simples de inteligencia nos llevará a un mejor entendimiento de la nuestra, y en general, de las formas de inteligencias más complejas.

Esta última consideración está determinada fundamentalmente por la forma en que Brooks entiende los procesos evolutivos de los organismos a través de la historia. El desarrollo de los microorganismos en otros más complejos es un conjunto de procesos que acontecen en largos periodos de tiempo. Por lo mismo, Brooks cree que cosas como el lenguaje o el comportamiento de resolución de problemas son mucho más fáciles de comprender cuando la esencia de ser y reaccionar están, por así decirlo, ya disponibles para el organismo. Se podría decir que dando este primer paso, lo otro es sólo cosa de tiempo (y de paciencia).

A partir de esto, Brooks ha creado algunos robots en base a los principios metodológicos ya supuestos²⁸. Estos robots lo han hecho llegar a una *conclusión inesperada (C)* y a una *hipótesis radical (H)*:

(C): Se ha observado que en niveles simples de inteligencia, las representaciones explícitas y modelos del mundo siempre resultan perjudiciales. Es mejor usar el mundo como su propio modelo.

(H): La representación es la unidad equivocada de abstracción en la construcción de las partes más voluminosas de los sistemas inteligentes.

Las razones que esgrime Brooks para llegar a **(C)** se encuentran en su mayor grado implícitas en lo que se llamó *problema del dinamismo*. Aun así, se hará una profundización al respecto en la siguiente sección.

²⁸ Una ampliación de estos principios así como una descripción de dos de los robots construidos por Brooks se encuentran al final de este capítulo

3.2.2 – Representaciones

Se ha intentado mostrar hasta ahora que la *abstracción* puede ser un arma peligrosa (Brooks, 1991b, p. 2) para los objetivos trazados. Por mucho tiempo se ha mantenido la opinión de que la clave de la inteligencia se encuentra en obtener la “*abstracción correcta*”. Sin embargo, lo que se hace al intentar buscar el *software* correcto para la computadora con el fin de que lleve a cabo una tarea específica, es hacer el trabajo por la máquina. Una máquina diseñada para comportarse como un organismo biológico debe contar con lo que llamamos anteriormente *principio de reactividad*. Es en este principio donde desaparece la línea divisoria entre percepción y cognición, creando así la posibilidad de que una máquina ejecute una conducta sólo por el hecho de estar percibiendo información del ambiente. Si suponemos que existe una *abstracción correcta*, no seremos capaces de superar la línea divisoria existente entre percepción y cognición. Al seguir confiando en esta distinción, obtendremos un errado entendimiento de los procesos de percepción de los organismos, ya que, continuaremos en la búsqueda del programa correcto. Como vemos, la adopción del principio de reactividad es importante tanto para la explicación de la conducta en ciencias cognitivas (el ejemplo de Uexküll mostró esto) como para el diseño de máquinas en la IA. Una máquina que no se basa en el principio de reactividad para su creación y confección, es una máquina cuyo marco está basado en la introspección del creador. Como dijimos anteriormente, lo relevante para el comportamiento se encuentra en el mundo, y se obtiene al poder percibir el ambiente de manera directa, no es posible obtener esta información sólo a través del “parecer del ingeniero”. Brooks usa en este respecto la terminología de Uexküll para referirse a este problema: le brindamos a las máquinas (y también a los animales en la explicación de su conducta) un *unwelt*²⁹(ambiente) basado puramente en el que creemos que es el nuestro. Pero hay dos equivocaciones en esta hipótesis:

²⁹ Entender aquí *unwelt* como se describió al comienzo de este capítulo: como la unión del mundo perceptivo y del mundo efectivo de un organismo.

- i) Lo que pudiera ser el *unwelt* de un organismo, sólo aquel organismo lo puede experimentar
- ii) Lo que nosotros creemos que es de hecho nuestro *unwelt* es una creencia basada en la abstracción de otras creencias. Podemos tener una correcta descripción de él observando la conducta de las personas, pero creer saber cuáles son los límites en los que se diferencian percepción y cognición (si es que existen) y después abstraerlos es un método peligroso.

La abstracción en la IA clásica, conlleva dos tipos de *representación* que parecen ser perjudiciales para el entendimiento de la inteligencia. Estas son el *modelo*, que funciona como *representación central* y los *símbolos* que son manipulados por los mecanismos internos del organismo. Estos símbolos son de alguna manera implicados por la representación central: es el programa usado de base y que actúa como marco general el que propone aquellas nociones preceptuales que tendrán *contenido*. Por decirlo de otra manera, los símbolos son *el umbral que determina si un input perceptual es o no un percepto*, dividen lo que tiene significado para el sistema de lo que no. De alguna manera esto ya se encuentra en la representación central, pero ciertamente a un nivel diferente.

3.2.3 – Cuatro aspectos fundamentales

De la ingeniería de sus máquinas, Brooks identifica 4 aspectos fundamentales. Estos aspectos son quizás el aporte filosófico más grande que hace Brooks al entendimiento de la inteligencia durante su investigación. El rol que cumplen estas características es el de dar el pie para la creación de una nueva *mitología* en nuestro entendimiento de la inteligencia y de la cognición. Hay que considerar que las características están enfocadas para servir de guía en la *construcción de robots basados en el comportamiento*, por lo que su relevancia en ciencias cognitivas es una de las apreciaciones que intento fundamentar en esta tesis.

I) Situacionalidad

“*El mundo mismo es su mejor modelo*” (Brooks, 1991a, p. 15) (Trad. Autor)

Este punto se ha tocado con anterioridad en lo que respecta a la importancia de la percepción directa en la teoría de la cognición de Brooks. Para que un robot pueda desarrollar tareas de la manera en lo que lo hace un organismo biológico, este debe estar situado en el mundo. Esto no quiere decir otra cosa que la máquina no debe tener ninguna representación general que funcione como su *modelo*. La máquina debe percibir el ambiente de manera directa, teniendo una experimentación continua del ambiente sin que haya procesos internos que medien esta información transformándola en otra. Como se intentaba explicar con el ejemplo de la naturaleza de los músculos en los postulados de Jakob Von Uexküll, un “operador de máquina” debe reaccionar ante todos los estímulos del ambiente traduciéndolo en un comportamiento que está determinado por la intensidad del estímulo. Esto es importante por dos razones: (i) le otorga continuidad al organismo en su conducta y (ii) lo lleva más cerca de la naturaleza de los organismos biológicos. La inteligencia como fenómeno es algo que ocurre en la medida que el agente, de la manera más eficiente posible y ante el constante bombardeo de estímulos del ambiente, logra realizar una actividad. Esto no puede ocurrir si es que el organismo no está de hecho lidiando con el mundo real. La crítica va dirigida en este sentido tanto al modelaje no real como al testeo de robots en *ambientes controlados* los cuales se alejan de las dificultades que impone desenvolverse en un mundo al completo. Creer que es mejor darle a un robot un mundo más pequeño con el que lidiar para maximizar su desempeño, y luego, paulatinamente, ir ampliando su espectro de contacto con el ambiente, es no darse cuenta de la esencia de los problemas del *dinamismo* y la *relevancia*. Una máquina que no se encuentra inserta en el mundo real es una máquina que, para cualquier cambio dentro de su sistema de creencias que se quiera hacer, se deberá hacer una reestructuración total de las relaciones de esas creencias. Ese es exactamente el trabajo que le corresponde a la máquina hacer, y no a nosotros.

En este punto puede establecerse la diferencia existente entre la percepción de *tokens*, característica del paradigma clásico, y la percepción de objetos o propiedades funcionales, que es, por otra parte, lo que pretende Brooks con su robótica. Para ilustrar esta diferencia, pondré el ejemplo de la creación un robot cuya función es, por decirlo así, evitar ser mordido por un perro. La versión clásica de este robot haría que éste *percibiera con conceptos* lo que es un perro, es decir, tendría en su memoria una categoría bajo la cual debería caer el input perceptual para ser percibido como perro. Podemos imaginar ahora la cantidad de posibles errores de categorización que podría tener eso. El robot debería tener una definición sumamente específica del concepto perro, de manera que no lo confunda con gato, o con mesa, o con perro de peluche. Por otro lado, el proceso de identificación del perro se lleva a cabo de manera conjunta con la necesidad de evitar ser mordido por el perro. Si el ejemplar de perro en el experimento es problemático (un perro poddle fácilmente podría ser mal categorizado como perro de peluche) la computación sería demasiado lenta para que el robot llevara a cabo su objetivo. Los ejemplares de perro que deberían ser usados para el testeo del robot deberían ser cuidadosamente seleccionados, y esto atentaría con el objetivo de lograr que el robot se desenvuelva en el mundo real. Es también problemático pensar que el robot sería programado específicamente para escapar sólo de la mordida de los perros. Podría ser el caso de que esta función fuera sólo una manera de mantener al robot a salvo de eventos que pudieran perjudicar su integridad física. Si es así, el robot no podría evitar una mordida de un león porque este no cae bajo el concepto “perro”. Luego, si nos atrevemos a incluir a todos los animales que pudieran morderlo o bien la explosión computacional o bien el lentísimo, y por lo tanto inútil proceso de reconocimiento del peligro, serían inevitables.

En cambio, lo que Brooks propone, es que el error está en creer que se necesita categorizar algo para llevar a cabo la tarea eficientemente. La máquina no debe percibir “perro”, debe percibir “aquello-que-puede-o-va-a-morderme”. Este *mecanismo de comportamiento* funciona como la descripción del músculo hecha por Uexküll: reacciona ante la intensidad del input transformándolo en comportamiento,

no categorizándolo. Esto es el corazón de lo que Brooks llama *situacionalidad*, es decir, la capacidad reactiva que tiene la criatura. Una naturaleza reactiva de esta clase no posee los problemas del *dinamismo* y de la *relevancia* expuestos anteriormente. Un organismo de esta naturaleza es *robusto*, no cuenta con leyes rígidas que al ser violadas le impidan extraer información del ambiente. El organismo extrae información del estímulo porque éste es esencialmente información, no el *token* de una clase previamente almacenada. Es difícil creer que un organismo debe extraer clases del ambiente cuando este es variable y múltiple. La situacionalidad implica leyes suaves que permitan dar cuenta de esa variabilidad múltiple.

II) Corporización

“El mundo necesita un camino de ingreso” (Brooks, 1991a, p. 16) (Trad. Autor)

Una característica como la situacionalidad no puede darse en un agente que no cuente con un cuerpo por el cual pueda obtener información acerca del ambiente. Por mucho tiempo se ha pensado que hay un nivel en que la cognición humana funciona de manera solipsista: existe una parte de nuestra cognición que se constituye de manipulación de símbolos los cuales no refieren directamente a nada del mundo físico. Esto también va de la mano con la pregunta recurrente en filosofía de si puede existir una mente sin cuerpo. Sin embargo, la segunda consideración surge ante la creencia no obligatoria de que la mente es algo distinto del cuerpo. Como he tratado de señalar, a través de las investigaciones de Uexküll y Brooks fundamentalmente, adscribir complejos procesos mentales para la explicación de la conducta de los organismos es quizás innecesario. Si es así, entonces cabe preguntarse qué rol puede jugar aún la mente en la explicación de la conducta. En el apartado subsiguiente, que corresponde a la *emergencia*, seguiré con este problema.

Hasta ahora se han dado fuertes razones para pensar que es posible entender al cuerpo como una parte fundamental de la mente, y que los experimentos mentales de “cerebros en bateas” son sólo ilusiones o ficciones que no argumentan a favor de la

existencia de una mente inteligente sin un cuerpo. Es parte de la tesis defendida en este trabajo aseverar que ante la evidencia mostrada por la investigación de Brooks, la manipulación simbólica no es la base de una conducta inteligente, inclusive, quizás ni siquiera es una parte de ella. Al contrario, la base de todo comportamiento inteligente es el nexo inmediato generado entre el agente y el ambiente. Este nexo es sólo posible si el agente posee un cuerpo que cuente con sensores y actuadores que le permitan comportarse de manera *dinámica* con el mundo.

III) Inteligencia

“La inteligencia está determinada por las dinámicas de la interacción con el mundo”
(Brooks, 1991a, p. 16) (Trad. Autor)

Esto ya se ha mencionado constantemente en este trabajo: la base de todo comportamiento inteligente no son los procesos complejos, sino los más simples:

“...las cosas ‘simples’ que tienen que ver con percepción y movilidad en un ambiente dinámico le tomaron a la evolución mucho más tiempo de perfeccionamiento [que los procesos complejos], y todas esas capacidades son una base necesaria para el intelecto de ‘alto nivel’.” (Brooks, 1991a, p. 16) (Trad. Autor)

Los procesos que nosotros creemos de menor complejidad que aquellos que son característicos de los seres humanos (como el lenguaje, la resolución de problemas, etc.) son de hecho necesarios para que se den los otros. La evolución no consiste en un proceso que se da de manera que hay organismos que han sido *intrínsecamente complejos* durante toda su historia evolutiva. Si vamos a basar nuestra metodología de investigación en la importancia de considerar cómo se ha llevado a cabo la evolución en los organismos biológicos, no podemos pensar que los procesos considerados más complejos pueden ser entendidos aparte de los que surgieron primero y que se manifiestan de mayor manera en casi todos los organismos.

IV) Emergencia

“La inteligencia está en el ojo del observador” (Brooks, 1991a, p. 16) (Trad. Autor)

Este es, para los fines de esta investigación, el punto más importante en la hipótesis teórica – filosófica de Brooks. En el primer capítulo de esta tesis convenimos en que la inteligencia debe ser investigada como un fenómeno extendible a cualquier organismo capaz de comportarse. Estos comportamientos deben estar sujetos a la necesidad del ambiente y del agente. Es ahí donde podemos establecer la diferencia de los comportamientos que tienen la roca que no juega ajedrez y el no jugador de ajedrez del ejemplo de la p. 11. Uno de los comportamientos, no es un comportamiento propiamente tal, porque no está motivado por las necesidad ni de él ni del ambiente (el de la roca), en cambio el comportamiento del no jugador de ajedrez está determinado por su incapacidad de jugar ajedrez en el momento convenido, pudiendo ser explicada la negativa a jugar ajedrez como un comportamiento desencadenado por una necesidad establecida por el ambiente. Ahora bien, la inteligencia se da en ese mismo respecto, sin necesidad de recurrir a los mecanismos por los cuáles aquellas entidades pudieran haber determinado su conducta. Lo que observamos es (en el caso del no jugador de ajedrez) un organismo que se adecúa a la situación dada por el ambiente, utilizando las *posibilidades de acción* que le brinda su *unwelt*. Dijimos antes que la inteligencia puede ser entendida como un comportamiento exitoso para el organismo ante las eventualidades del ambiente, y que lo que sea un “comportamiento exitoso” lo determina el mundo (ambiente) mismo. Esto es verdad sólo en parte, puesto que la inteligencia no es algo que el ambiente adscriba a los organismos o a sus conductas, sino que esto es hecho por un *testigo*. Los que determinan finalmente si los comportamientos efectuados por algo o alguien caen bajo lo que consideramos inteligente son aquellos que al parecer se comportan de manera inteligente con bastante frecuencia (es decir, los seres humanos). La inteligencia, en este sentido, “emerge de la interacción de los componentes del sistema” (Brooks, 1991a, p. 16) (Trad. Autor); pero esto no hay que entenderlo en el sentido tradicional de lo que conlleva el término *emergente*. La

inteligencia no juega ningún rol en la explicación de la conducta, es, de hecho, el *explanandum* de la investigación de Brooks (al igual que el de este trabajo). Lo que sea la inteligencia emerge de la interacción de los componentes del sistema, pero no en el sistema, sino en el testigo: “La inteligencia sólo puede ser determinada por el comportamiento total del sistema y cómo ese comportamiento aparece en relación con el ambiente.” (Brooks, 1991a, p. 16) (Trad. Autor)

3.2.4 – Inteligencia y cognición

Podemos establecer ahora una relación más concreta entre la *cognición* y la *inteligencia*. Ya se describieron ambos fenómenos como ocurriendo en el ojo del observador, el cual ve al organismo que se comporta de manera inteligente. La relación conceptual entre los elementos explicativos introducidos en este trabajo podría darse de la siguiente manera. Un observador contempla la conducta “x” de un organismo. Esta conducta cumple con los criterios de *diversidad* y *concordancia* que son característicos de las conductas inteligentes, por lo que el observador califica el fenómeno observado como muestra de un comportamiento inteligente. El observador cree que la conducta “x” es de naturaleza compleja, por lo que la explica a través de mecanismos internos al agente que dan lugar a lo que llama *cognición*. La *cognición* y la *inteligencia* son así fenómenos que son adscritos por un observador: el primero explica el segundo.

Por último, ubicando a lo *mental* dentro de esta cadena conceptual, si creemos que la *cognición*, al dar lugar a comportamientos complejos, no es necesaria para la explicación de comportamientos más simples de organismos más simples, podemos también pensar que el fenómeno de lo cognitivo se lleva a cabo en un reino diferente al puramente físico y reactivo, pudiendo establecerse aquí un vínculo con la *mente*. De esta manera, lo *cognitivo*, lo *inteligente* y lo *mental*, están determinados entre sí. Si adoptamos una explicación de la conducta basada en la definición de inteligencia defendida en esta tesis, podemos simplemente prescindir de conceptos como *cognición* y *mente*. Podemos prescindir del primero porque al estar éste fuertemente

ligado a la explicación clásica de la conducta, puede estar sujeto a confusiones; al fin y al cabo lo que se intenta señalar con él en la explicación corporizada es, simplemente, la conjunción de los procesos de percepción y acción. El segundo es prescindible por su clara inoperancia en la explicación propuesta hasta ahora; si existe algo así como lo mental resulta altamente irrelevante para la explicación de la conducta.

Entonces, ¿por qué deberíamos trasladar la explicación de la inteligencia desde los mecanismos internos del individuo a la complejidad de ambiente?

- 1) Como se describió al comienzo del capítulo, la manera en que comúnmente entendemos las nociones de lo cognitivo, lo inteligente y lo mental es sólo una manera de comprender la conducta de los organismos, y esta manera no está sujeta a ningún tipo de necesidad. Podemos, estipulando pre - conceptos coherentes y económicos, adscribir libremente a otro tipo de explicación.
- 2) La manera anterior de comprender estos fenómenos no nos lleva a la explicación pretendida en este trabajo como objetivo de las ciencias cognitivas y la IA. Evidentemente la explicación de la cognición corporizada tiene también reparos posibles, además que (1) también se le aplica a ella.

A continuación se expondrá la metodología con que Brooks pretende llevar a cabo este marco teórico en sus robots. Como esta tesis pretende ser relevante en la investigación de la inteligencia tanto para las ciencias cognitivas como para la IA, creo que la descripción de la arquitectura de las que Brooks llama *criaturas*, es muy importante para el develamiento de nuevas evidencias para la comprensión de lo inteligente.

3.2.5 - Criaturas

El objetivo de Rodney Brooks es: "...crear agentes móviles completamente autónomos que coexistan en el mundo con los humanos, y sean vistos por esos humanos como seres inteligentes a su manera" (Brooks, 1991b, p. 4) (Trad. Autor)

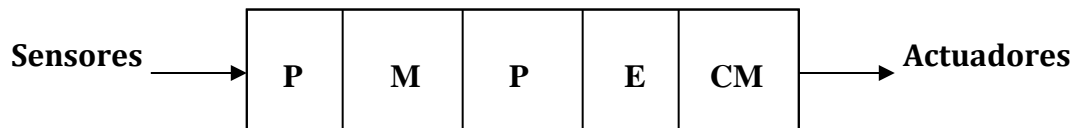
Esto es a lo que Brooks llama *criaturas*. Estas deben ser *dinámicas*, deben acoplarse de manera efectiva con el ambiente y sus cambios, y *robustas*, que cambios menores en su ambiente no lleven al colapso total del robot. Además, estas criaturas deben poder mantener metas múltiples e intentar llevarlas a cabo de manera paralela. La esencia de la criatura es que esta sea una colección de competencias de conductas que intentan llevarse a cabo de manera simultánea. Cabe recordar que la esencia de esta precondition es que la *inteligencia de alto nivel* surge de la interacción de procesos de percepción-acción más simples que son propios y fundamentales de la inteligencia. La hipótesis aquí es que aquellas conductas más complejas emergen de una especie de *caos conductual*, el cual si bien no es necesario para dar lugar a la inteligencia, la experimentación ha mostrado que conductas en organismos (artificiales) que parecen complejas son el producto de una colección de conductas más simples (Brooks, 1985).

3.2.6 - Hipótesis del enraizamiento físico

Ya habiendo establecido las características que deberían tener las criaturas, ¿cómo es posible llevar esas características a cabo en la inteligencia artificial? Rodney Brooks comienza estableciendo una hipótesis metodológica diferente a la expuesta por Newell y Simon. Esta hipótesis tiene el nombre de *hipótesis del enraizamiento (o fundamento) físico*, y consiste en que cualquier tipo de representación que el robot use, no puede ser un símbolo, es decir, no puede ser la representación de una representación, debe estar fundada en el mundo físico. Los alcances de esto son que el robot nunca estaría lidiando con un *concepto* sino simplemente *percibiendo para comportarse*. La hipótesis hace que el robot no deba interpretar la información, es decir, no manipula información almacenada que pueda combinar con la información perceptual para luego deliberar qué es lo que debe hacer. El robot no posee más información acerca del percepto por manipularlo junto a otras representaciones, la información que tiene del percepto es la misma cuando entra y cuando se comporta. Lo que ocurre dentro de él no es interpretación, sino el traspaso de una descripción.

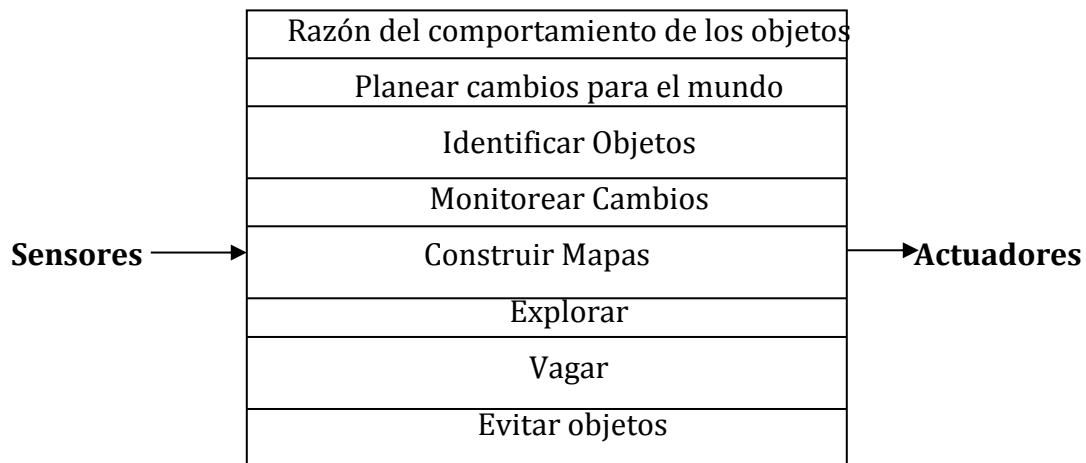
Para lograr primeramente esto debemos cambiar la tradicional estructura horizontal de las máquinas de la IA clásica por otra que no involucre un *procesador central* que regule los procesos internos del robot. La razón es la expuesta más arriba, la ingeniería del robot debe ser dispuesta de manera que éste pueda ejecutar varios comportamientos a la vez sin necesidad de que estos sean regulados. En las siguientes dos figuras se puede ver el contraste entre la descomposición de un sistema de control tradicional y la que es usada por Brooks en sus robots.

Descomposición clásica del sistema de control de un robot móvil en módulos funcionales:



P: Percepción **M:** Modelaje **P:** Planeamiento **E:** Ejecución de Tarea **CM:** Control Motor

Descomposición del sistema de control de un robot móvil basada en comportamientos realizadores de tareas:



Este tipo de descomposición (el de Brooks), permite que el robot pueda ejecutar comportamientos diferentes de manera simultánea sin poseer un procesador central

que regule aquellos comportamientos. La forma para poder llevar a cabo esto es una arquitectura computacional conocida como *arquitectura de la subsunción*.

3.2.7 – Arquitectura de la Subsunción

Una criatura es una competencia de diferentes actividades. Estas actividades son patrones de interacciones con el mundo que en el robot tienen forma de *capas*. Una capa en este sentido es una actividad específica, tiene la forma de un algoritmo simple y su rol es llevar a cabo la conducta programada si es que es necesario. Estas capas son de naturaleza *incremental*: es posible que a un robot puedan añadirse nuevas capas en el futuro incluso después de ya testeadas las anteriores. Esto, ya que el robot carece de un procesador central que regule el funcionamiento de las capas. En este sentido, las capas son altamente independientes unas de otras (pero no del todo, esto se explicará a continuación). Existen capas de orden superior capaces de inhibir a otras por algún periodo de tiempo si es preciso, pero las capas de niveles inferiores no tienen esta capacidad y pueden ser inhibidas por otras pero ellas no tienen incidencia sobre ninguna.

Cada capa consiste en una *red de máquinas de estados finitos aumentables* (desde ahora MEFA) las cuales cumplen el rol de traspasar la información requerida para así hacer posible la conducta del robot. Las máquinas procesan un código computacional cualquiera (en el caso de las máquinas Brooks, LISP) y la comunicación entre ellas se da a cabo a través de canales de input y output. Este proceso incluye, evidentemente, representaciones, pero no del tipo de los robots de la IA clásica. Estas representaciones, como dijimos antes, sólo comunican una descripción para llevar a cabo una conducta. Es decir, son de la misma naturaleza que el impulso nervioso que viaja desde un brazo al cerebro portando información acerca del golpe recibido en este. Una MEFA consiste en 4 partes:

-Máquina de estados finitos: Consiste en una máquina de estados finitos normal, la cual posee 4 posibles estados: 1) Enviar un mensaje al puerto designado y luego

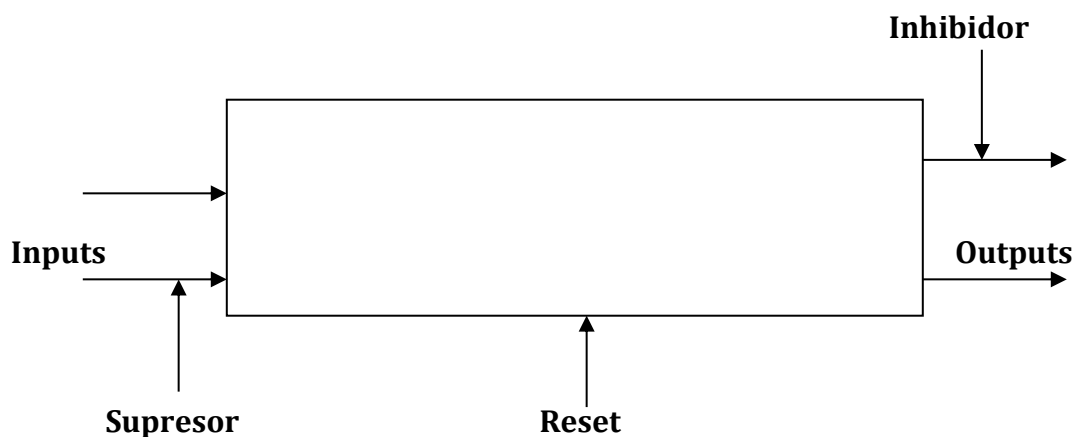
cambiar a otro estado 2) Testear el valor de una función periférica³⁰ (mensaje) y a partir de ello cambiar a otro estado. 3) Computar una función periférica y luego cambiar a otro estado (esta función queda escrita en los registros de las otras máquinas) 4) Esperar en paralelo por algún evento y cuando este ocurre cambiar a otro estado.

- **Registros:** Los registros son escritos por las otras máquinas y, por lo general, se escriben en ellos funciones periféricas las cuales asignan una condición para cambiar a otro estado. Esta información puede ser traspasada a otras máquinas.

- **Relojes:** En la MEFA también hay relojes [*timers*] los cuales actúan como cronómetros. Asignan un tiempo para detonar un cambio a un estado específico.

La razón de que las máquinas sean aumentables es la misma de porqué las capas son incrementales: es posible añadir nuevas MEFAs sin perjudicar el funcionamiento de la capa en cuestión.

La manera en que ciertas capas pueden ser inhibidas es a través de los canales de inhibición y supresión.



Los canales de supresión actúan sobre el input de la MEFA, impidiendo que cualquier mensaje que vaya por la conexión llegue a ella. Los canales de inhibición actúan de la

³⁰ Una función periférica no es más que el cálculo de la información recibida por varios canales de input. Es una manera de saber el estado de las otras máquinas.

misma manera pero en los canales de salida (outputs) de la MEFA. Es decir, anulan cualquier mensaje enviado desde ésta. Ahora, estos canales tienen dos maneras de ejecutar sus objetivos. El canal supresor o inhibidor puede actuar sobre el canal correspondiente enviándole un mensaje que impida el arribo o la salida de mensajes por un periodo de tiempo establecido de manera arbitraria por el programa inhibidor o supresor. La otra manera es que los canales actúen enviando mensajes durante todo el tiempo que la inhibición o la supresión sea necesaria. Esta última manera ha resultado ser mejor según Brooks, puesto que permite una mejor coordinación entre las capas, evitando la arbitrariedad de programar un tiempo para llevar a cabo el acto inhibitorio (Brooks, 1990, p. 4). Lo particular de este sistema de subsunción, es que permite cierta jerarquía entre las capas pero no un control completo de una sobre otra (o de una sobre todas). Inhibir o suprimir una capa en el sistema de Brooks es una cuestión física, y no de programa. La razón que tiene una capa para incidir sobre otra es porque hay dos comportamientos que físicamente, no pueden llevarse ambos a cabo de manera simultánea. Es sensato pensar que si un robot tiene las actividades de “vagar libremente” y “ver un lugar e intentar alcanzarlo”, ha de existir un mecanismo que permita a ambas convivir pero sin que ambas se lleven a cabo simultáneamente. No creo que una persona pueda de hecho realizar aquellas dos conductas al mismo tiempo. Esto no implica que esa sea la manera en que una persona especifica su conducta, pero tampoco desestima la posibilidad de que así sea. Por otra parte, si se logra conectar las capas de un robot de manera que su comportamiento pareciera ser el de un animal o inclusive una persona, poco importará si esa es la manera real en que un ser biológico determina su conducta. Lo que se quiere conseguir a través de este tipo de robótica y cómo se consigue son dos cosas diferentes. En la metodología adoptada prima lo primero por sobre lo segundo (de hecho, esa es la metodología de Turing). Existe un canal adicional conocido como *reset*, el que permite a la máquina volver a un estado inicial si es que es conveniente.

3.2.8 – Dos instancias de la metodología

Allen

Allen es un robot creado por el departamento de ingeniería de IA del MIT, en el tiempo que éste era presidido por Rodney Brooks. Este robot deambula libre por los pasillos del MIT en razón de la máxima de que toda criatura debe ser testada en el mundo real. Allen se mueve a través de la odometría, es decir, calcula su posición a través del número y el tipo de vueltas que dan sus ruedas. Como mecanismo de percepción utiliza un generador de ondas de sonar y un captor de las mismas. Ésta criatura está construida en base a la arquitectura recién expuesta, es decir, es un robot que consta de tres capas. Cada una de ellas está conectada a los receptores y es capaz de ejecutar un comportamiento, es decir, cada capa está encargada de una actividad particular.

La capa inferior evita que Allen choque con objetos, tanto estáticos como dinámicos, la segunda capa hace que Allen vague de manera aleatoria por el terreno, y la tercera capa permite a Allen escoger un lugar objetivo e intentar alcanzarlo. El funcionamiento de las capas es el siguiente: cuando Allen se encuentra sólo en la mitad del campus MIT, su segunda capa le permite vagar libremente hasta que la capa de alcanzar objetivo inhibe a la segunda y pone a Allen en proceso de ir a un lugar específico. Mientras Allen vague o intenta alcanzar un lugar, la primera capa evitará los obstáculos que vaya encontrando en el camino. Esto lo hará sin tener que interferir de ninguna manera con los procesos llevados a cabo por las otras capas. Lo que intenta mostrar Allen es que es posible poseer dos “productores de actividades” paralelos los que no convergen en un mismo comportamiento; esto hace que desde fuera pareciese que el robot evita obstáculos en pos de un objetivo ulterior, pero esto no es así. Lo que se lleva a cabo dentro de Allen es una competencia de conductas, aunque, dada la naturaleza de las actividades, la probabilidad de que se desencadene algún tipo de caos es mínima. En conclusión, cada capa es un algoritmo diferente que ejecuta una habilidad diferente sin la necesidad de un objetivo común.

Herbert

Herbert es un robot más complejo que Allen. Posee una mayor cantidad de sensores además de un aún más complejo sistema de procesadores internos. Estos procesadores son más lentos que los de otros robots creados por el equipo de Brooks, pero esto hace que use menos energía eléctrica, permitiéndole mayor tiempo de autonomía antes de cargar sus baterías. El sistema general que usa Herbert sigue siendo el mismo que funciona en Allen. Sin embargo, este posee dos diferencias que lo hacen especial. Por una parte, en lugar de usar un emisor de ondas de sonar, Herbert utiliza un sistema delineador de luz láser, capaz de brindarle información de profundidad en 3D acerca del entorno cada segundo. Segundo, Herbert no posee canales de inhibición entre sus capas, es decir, en Herbert realmente cada capa funciona de manera autónoma. Las actividades de Herbert también incluyen evitar obstáculos y vagar, pero en lugar de alcanzar lugares – objetivo, lo que hace Herbert es buscar y coleccionar latas de aluminio de los escritorios en las oficinas. Herbert posee un brazo mecánico cuidadosamente diseñado para poder alcanzar latas en lugares que puedan ser dificultosos. La manera en que Herbert actúa es la siguiente: Cuando llega a una habitación, Herbert colecciona información que le permite saber los objetos que hay sobre los escritorios. Si identifica una lata, Herbert tomará posición para poder alcanzarla. Lo que detona el movimiento del brazo, sin embargo, no es la información 3D acerca de la lata, sino el cese de movimiento de las ruedas. Luego de tener éxito en alcanzar la lata, Herbert la lleva a su punto de partida. Este robot mostró ser eficaz para la tarea que fue diseñado, demostrando un alto porcentaje de éxito en el reconocimiento de las latas, así como mostrándose naturalmente oportunista ante las ocasiones que se le presentaban.

Conclusión

La investigación llevada a cabo en este trabajo ha pretendido, al menos, llegar a sembrar la discusión acerca de cómo entendemos fenómenos como la inteligencia y la conducta. Si además de eso, el trabajo ha logrado convencer de que la cognición corporizada es una propuesta que debe ser tomada en serio de aquí en adelante tanto en ciencias cognitivas como en la inteligencia artificial, el objetivo de esta tesis está aún mejor cumplido.

A lo largo de este trabajo se ha expuesto cómo es que pareciera que la inteligencia es más bien, algo de lo que estamos seguros al adscribir a los demás, que algo que estamos dispuestos a explicar (1). También se mostró cómo se ha desarrollado la investigación de las ciencias cognitivas y la IA acerca de los fenómenos mencionados en los últimos 60 años, y cuáles son los problemas que el paradigma clásico acarrea (2). Por último, se intentó exponer las bases para una nueva comprensión de la conducta y lo inteligente a través de la propuesta de algunos investigadores del siglo pasado (Gibson, Uexküll) como también a través de las *criaturas* diseñadas por Brooks (3).

La cognición corporizada ha sido exitosa en la creación de robots con las características que ella misma ha especificado como esenciales para que acontezca la inteligencia. Entender la cognición como la conjunción de la percepción y la acción en el ojo del testigo ha permitido superar los problemas de la *relevancia* y el *dinamismo* expuestos en este trabajo. Por otra parte, esta perspectiva de explicación para la conducta permite dejar en evidencia que muchas de las concepciones que se mantienen hoy en día en ciencias cognitivas sólo han sido argumentadas desde la *tenacidad* y no desde la experimentación científica.

Hoy en día es posible leer las descripciones o ver los videos de cientos de robots que funcionan de manera autónoma y que se basan en los principios reactivos de la filosofía de Brooks (pero no todos en la arquitectura de la subsunción). Aquel éxito

nos muestra lo siguiente: no es descabellado pensar que procesos como los que se llevan a cabo en estas máquinas puedan en un futuro lejano dar lugar a organismos capaces de comportarse de una manera que alguien no pueda distinguirlos de las personas, pasando así el test de Turing. El camino para llegar a esto es por ahora, inmenso. Las arquitecturas basadas en la filosofía de la corporización no han podido aún dar cuenta de robots que posean conductas más complejas que evidencien un gran avance en nuestro acercamiento hacia la inteligencia humana. Inclusive Brooks (Brooks 1991b, pp. 10 – 11) se pregunta si estas máquinas podrán realizar procesos más complejos como el aprendizaje, al modo que aprenden los humanos³¹. Tampoco se sabe cuántas *capas* es posible acoplar antes de que el caos sea tan grande que sea necesario implementar algún tipo de *procesador central*. Estos y muchos otros retos aparecen en el horizonte de, no sólo la cognición corporizada, sino de todos los enfoques alternativos y anti – representacionalistas. Esto no desestima la importancia que tiene la investigación de los programas alternativos, inclusive si en el futuro todos llegasen a conclusiones falsas, porque por ahora: “Violar el santuario de la presunción y la ignorancia ya es, supongo, prestar algún servicio al entendimiento humano”.³²

No hay nada dicho en la investigación de las ciencias cognitivas y la inteligencia artificial. Los enfoques alternativos al paradigma clásico han mostrado tener una explicación más satisfactoria de las conductas reactivas o simples de los organismos. El enfoque clásico, a pesar de las críticas vertidas en este trabajo, ha logrado grandes avances en materia computacional, consiguiendo una mejor comprensión de los procesos abstractos, pudiendo también así mejorar de gran manera la tecnología de la que disponemos hoy en día. Lo que nos haga estimar un paradigma por sobre otro será el éxito de aquellos en el objetivo de la comprensión de lo inteligente. Quizás en el futuro algún tipo de unión pudiera dar paso a una explicación nueva acerca de la

³¹ Se han podido programar capas de aprendizaje a la manera en que una abeja aprende cuáles son las flores que debe polinizar y cuáles no (Brooks, 1991b). Este es el tipo de acercamiento al que Brooks apunta: entender el funcionamiento de las inteligencias más simples nos ayudará paulatinamente a entender las más complejas.

³² Locke, J., (1994), “Ensayo sobre el entendimiento humano”. Madrid: Fondo de cultura económica

conducta. Esto por ahora, sin embargo, parece lejano dadas las naturalezas incompatibles de ambos programas.

Confío en que los argumentos y las visiones expuestas en este escrito hayan logrado el objetivo del mismo: convencer de que la inteligencia entendida como aquello que distingue a los organismos biológicos es un método más que válido para la explicación del fenómeno de la conducta.

Lista de Referencias

- Anderson, M. (2003). Embodied Cognition: A field Guide. *Artificial Intelligence*(149), 91 - 130.
- Aristóteles. (18 de Enero de 2013). *The History of Animals*. Obtenido de Ebooks Adelaide: <http://ebooks.adelaide.edu.au/a/aristotle/history/>
- Brooks, R. (1985). A robust layered control system for a mobile robot. *Artificial Intelligence Memo*(864), 1 - 25.
- Brooks, R. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*(6), 3 - 15.
- Brooks, R. (1991a). Intelligence without Reason. *Computer and Thought, IJCAI - 91*, 1 - 27.
- Brooks, R. (1991b). Intelligence without Representation. *Artificial Intelligence*(47), 139 - 159.
- Brooks, R. (1999). *Cambrian Intelligence*. Cambridge: MIT Press.
- Calvo, P., & Gomila, T. (2008). Directions for an embodied cognitive science: Toward an Integrated Approach. En P. Calvo, & T. Gomila, *Handbook of Cognitive Science: An Embodied Approach* (págs. 1 - 25). San Diego: Elsevier.
- Descartes, R. (1937). *Discurso del Método y Meditaciones Metafísicas*. Madrid: Espasa Calpe.
- Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. New York: Psychology Press.
- Moravec, H. (18 de Enero de 2013). *Locomotion, Vision and Intelligence (1985)*. Obtenido de Hans Moravec, mobile robots since 1963: <http://www.frc.ri.cmu.edu/~hpm/project.archive/robot.papers/1983/mit.txt>
- Newell, A., & Simon, H. A. (2006). Computer Science as Empirical Enquiry: Symbols and Search. En J. L. Bermudez, *Philosophy of Psychology, contemporary readings* (págs. 407 - 432). New York: Routledge.
- Pfeifer, R., & Bongard, J. (2007). *How the body changes the way we think*. Cambridge: MIT Press.
- Rowlands, M. (2004). *The Body in Mind*. Cambridge: Cambridge University Press.

- Rumelhart, D. (1997). The Architecture of Mind: A Connectionist Approach. En J. Haugeland, *Mind Design II* (págs. 205 - 232). Cambridge: MIT Press.
- Smolensky, P. (1988). On the Proper Treatment of Connectionism. *Behavioral and Brain Sciences*(11), 1 - 23.
- Tienson, J. (1995). Una introducción al conexionismo. En E. Rabossi, *Filosofía de la mente y Ciencia Cognitiva* (E. Rabossi, Trad., págs. 359 - 380). Barcelona: Paidós.
- Turing, A. (1997). Computer Machinery and Intelligence. En J. Haugeland, *Mind Design II* (págs. 29 - 56). Cambridge: MIT Press.
- Turing, A. (2004). On computable numbers, with an application to the Entscheidungsproblem (1936). En J. Copeland, *The Essential Turing* (págs. 58 - 90). Oxford: Oxford University Press.
- Uexküll, J. V. (2010). *A Foray into the World of Animals and Humans*. Minneapolis: University of Minnesota Press.
- Wittgenstein, L. (1999). *Philosophical Investigations*. (G. E. Anscombe, Trad.) Oxford: Blackwell Publishers.