



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

INSTITUTO DE INVESTIGACIONES EN
MATEMÁTICAS APLICADAS Y EN SISTEMAS

MODELO DE ROBÓTICA EVOLUTIVA PARA EL
RECONOCIMIENTO EXPLÍCITO DE
AGENCIALIDAD

T E S I S

QUE PARA OPTAR POR EL GRADO DE:
Maestra en Ciencia e Ingeniería de la Computación

PRESENTA:
Leticia Cruz Bárcenas

TUTOR:
Dr. Tom Froese

Ciudad de México, 2017

Agradecimientos

En primer lugar agradezco a mi tutor por su paciencia, generosidad, ayuda, conocimiento, disponibilidad, pensamiento crítico. Por haberme motivado desde el comienzo de este proyecto de tesis, él fue una parte fundamental en las ideas presentadas aquí. Gracias por haberme mostrado las puertas de una nueva perspectiva que no era conocida para mí y que ahora me apasiona muchísimo. Al comité de sinodales: Dra. Katya Rodríguez, Dra. Wendy Aguilar, Dr. Mario Siqueiros y el Dr. Carlos Gershenson. Gracias por su valiosa e importante retroalimentación, enriqueciendo esta tesis con su experiencia y conocimientos.

A mis padres, por el amor recibido, la dedicación y la paciencia desde mis primeras muestras de vida. A mis hermanas y mi sobrino, que en conjunto son una familia que me ha dado soporte y ha estado en muchos momentos. Este proyecto de vida se los dedico a ustedes.

A mi amigo Jair, que estuvo desde el inicio de este viaje. Fue simplemente genial haber reído tantísimo con él, a pesar de que algunos momentos fueron de bastante estrés hubieron muchos más alegres. Gracias por las pláticas y por transmitirme tu forma de ser tan relajada. Todo eso es invaluable.

En la vida hay veces que nos encontramos seres especiales, alguien que cambia nuestra vida con sólo ser parte de ella. Alguien que nos hace creer que hay algo muy bueno y muy bonito en el mundo. A Marina, persona que se volvió un parteaguas en mi vida.

A la Universidad Nacional Autónoma de México, fue una experiencia verdaderamente increíble el haber tenido la oportunidad de estudiar aquí. Me encanta la sensación de estar en Ciudad Universitaria y siento un profundo orgullo y respeto por esta institución que se ha ganado a pulso el ser la máxima casa de estudios.

Finalmente, agradezco a CONACYT por el apoyo otorgado al posgrado, por los recursos brindados y que a pesar de las cuestiones políticas, continúa impulsando la ciencia y tecnología en México.

Declaración de autenticidad

Por la presente declaro que, salvo cuando se haga referencia específica al trabajo de otras personas, el contenido de esta tesis es original y no se ha presentado total o parcialmente para su consideración para cualquier otro título o grado en esta o cualquier otra Universidad. Esta tesis es resultado de mi propio trabajo y no incluye nada que sea el resultado de algún trabajo realizado en colaboración, salvo que se indique específicamente en el texto.

Leticia Cruz Bárcenas. Ciudad de México, 2017.

Índice general

1. Introducción	1
1.1. Objetivo general	3
1.1.1. Objetivos específicos	3
1.2. Metodología	3
1.3. Contribuciones	4
1.4. Estructura de la tesis	4
2. La ciencia cognitiva	7
2.1. Orígenes	7
2.1.1. Definición y disciplinas subyacentes	7
2.1.2. Dualismo y conductismo	9
2.2. La cibernética	11
2.2.1. Cibernética de Ashby	12
2.2.2. Cibernética de segundo-orden	13
2.3. Inteligencia artificial simbólica	14
2.3.1. Teoría computacional de la mente	14
2.3.2. Representación del conocimiento y computacionalismo	15
2.3.3. Argumentos y críticas sobre el representacionalismo y la teoría computacional de la mente	16
2.4. Enfoques alternativos	19
2.4.1. Conexionismo	20
2.4.2. Dinamismo	20
2.5. Retos para la ciencia cognitiva	21
3. Hacia una nueva ciencia cognitiva	23
3.1. La inteligencia artificial corporizada-embebida	23
3.2. Ciencia cognitiva enactivista	24
3.2.1. Autonomía	24
3.2.2. Identidad	25
3.2.3. Agencialidad	25
3.3. Dimensiones de la corporización	26
3.3.1. Autorregulación corporal	26
3.3.2. Acoplamiento sensomotor	26

3.3.3. Interacción intersubjetiva	27
4. Metodologías y herramientas de estudio	29
4.1. Robótica evolutiva	29
4.1.1. ¿Qué es?	30
4.2. Redes neuronales artificiales	31
4.2.1. Redes Neuronales Artificiales <i>feed-forward</i>	31
4.2.2. Redes Neuronales Recurrentes	32
4.3. Redes Neuronales Recurrentes de Tiempo Continuo	33
4.3.1. Descripción técnica de las CTRNN	34
4.3.2. Comportamiento de las CTRNN	35
4.4. Algoritmos genéticos	35
4.5. Teoría de sistemas dinámicos	36
4.5.1. Conceptos	36
4.5.2. El enfoque de los sistemas dinámicos	37
4.5.3. Sistemas cognitivos como sistemas dinámicos	38
5. Modelos Mínimos de Comportamiento Cognitivo	39
5.1. Cognición Social	39
5.2. El paradigma del cruce perceptual	40
5.2.1. Modelos sintéticos mínimos de cognición social	43
5.2.2. Modelos sintéticos de cruce perceptual	44
6. Modelo de robótica evolutiva para el reconocimiento explícito de agencialidad	47
6.1. Diseño	47
6.2. Procedimiento	49
6.3. Resultados	50
6.3.1. Resultados estadísticos	53
6.4. Análisis	57
6.5. Discusión	58
6.6. Conclusiones	60
7. Conclusiones	63
7.1. Trabajo a futuro	65
Bibliografía	67

Introducción

Desde sus inicios, el área de la computación y en particular la inteligencia artificial han sido de gran importancia en el desarrollo de muchas disciplinas y en la creación de otras. Gracias a la gran potencia que tienen sus herramientas algunos eventos que suceden en la naturaleza ahora pueden ser estudiados y comprendidos más fácilmente. Gran parte de las bases de la computación quedaron establecidas con el trabajo hecho por Alan Turing en el año de 1936 con la denominada *máquina de Turing*. La máquina de Turing es un modelo matemático para representar de manera lógica lo que es una computadora (Hodges, 2012). Con el surgimiento de este modelo, también se empezó a considerar la posibilidad de que la mente del ser humano pudiera comportarse de la misma manera. Esta idea adquirió fuerza en otras disciplinas enfocadas en el estudio del funcionamiento de la mente, como la psicología. Bajo este paradigma quedaron diseñados estudios, experimentos, investigaciones y aplicaciones en cuestiones dedicadas la mente y el cerebro. Sin embargo, conforme se iba avanzando en las investigaciones y el estudio de diversos fenómenos, comenzaron a surgir preguntas que ya no podían ser respondidas bajo ese esquema.

El estudio de la mente estuviera abordado principalmente por la filosofía y la psicología durante un periodo bastante amplio y después vendría a consolidarse una área que haría grandes aportaciones: la cibernética. La cibernética en el afán de crear sistemas adaptables al entorno puso en cuestionamiento los procesos y fundamentos sobre los cuales eran contruidos los sistemas a mediados del siglo XIX. Después tomaron fuerza enfoques basados en teoría de la información y representación del conocimiento, sin embargo éstos no ofrecían el marco de referencia necesario para crear sistemas verdaderamente autónomos y con otra serie de características importantes.

A lo largo de todos estos esfuerzos, comenzó a haber una marcada tendencia en implementar en esos sistemas características observadas en los humanos. Cuestiones como el aprendizaje, la memoria, la capacidad de adaptarse al cambio, la inteligencia, autonomía, etc., fueron de especial interés en diferentes áreas. La necesidad de replicar este comportamiento llevó al surgimiento de áreas como la inteligencia artificial y también al desarrollo de herramientas que pudieran hacer simulaciones de los mismos. Posteriormente surgiría la ciencia cognitiva en un esfuerzo conjunto de distintas disciplinas

1. INTRODUCCIÓN

por entender el fenómeno de la cognición.

La ciencia cognitiva se ha visto sumamente beneficiada por la inteligencia artificial, siendo una de sus aportaciones de esta última la creación de modelos para el estudio de la cognición social. La robótica evolutiva es una metodología de la inteligencia artificial que se ha encargado de generar herramientas para el estudio de comportamientos cognitivos. Básicamente, lo que se busca entender y explicar es cómo la coordinación que se da entre los individuos de un sistema se desarrolla para que se puedan alcanzar objetivos colectivos y que surja una auto-organización global. Además, los modelos sintéticos mínimos de cognición social propuestos por la robótica evolutiva han permitido reducir la complejidad de los sistemas de cognición social. Aparte, el análisis tiene dos características muy importantes: es sistemático y holístico (Froese et al., 2012).

En el estudio de la cognición social ha predominado el esquema en el que es suficiente con estudiar las propiedades y habilidades individuales para poder entender el comportamiento social. Además, se considera que los patrones de interacción son las entradas recibidas por módulos internos que poseen cada uno de los individuos de manera aislada. Lo anterior ha llevado frecuentemente a darle menor importancia a sus propiedades dinámicas. Algunas de las posibles razones por las que no se le da relevancia a este aspecto es porque se considera que los patrones de interacción son abstractos y complejos de entender. Como consecuencia, el resultado es que hasta el momento se tiene poco entendimiento acerca de cómo es la dinámica holística y en particular el papel que juegan las interacciones dentro de la cognición social. Por lo que es necesario que elementos cruciales como: los conceptos empleados, las herramientas utilizadas y los procedimientos desarrollados sean replanteados de una manera sistemática (De Jaegher et al., 2010). Se espera que nuevas herramientas conceptuales y analíticas faciliten el estudio de los mecanismos y factores que constituyen la cognición social.

Un experimento que ha sido de particular interés en el estudio de la cognición social ha sido el de Auvray et al. (2009). El experimento busca entender cómo es posible que los seres humanos puedan reconocer la presencia de otra entidad con intencionalidad, es decir, otra persona, del resto de los objetos del ambiente. Ese reconocimiento emerge de una dinámica presente en el cruce de las percepciones entre los dos individuos. Para estudiar este fenómeno fue diseñado un ambiente virtual en el que los participantes pudieran reconocer la presencia de la otra persona. Ese reconocimiento fue monitoreado a través del número de clicks. Este experimento permitió establecer lo que sería conocido el paradigma del cruce perceptual del cual se hablará en profundidad en los siguientes capítulos.

Los modelos de simulación del paradigma de cruce perceptual desarrollados hasta ahora no cuentan con el reconocimiento explícito por medio de clicks. En vez de eso utilizan la distancia entre los agentes como una medida de “reconocimiento” (cf. Di Paolo et al., 2008; Froese & Di Paolo, 2008, 2010, 2011). La razón de por la que se utilizó esa medida es debido a que simplifica y disminuye la dificultad de evolucionar agentes que tengan el comportamiento del click. Lo que han permitido estos modelos ha sido el estudio de las estrategias de los agentes artificiales para identificar al otro agente sobre el resto de los objetos, a través de la distancia relativa entre ellos. Sin embargo, no se

ha podido hacer un estudio completo de la verdadera estrategia que tienen los agentes al hacer un reconocimiento explícito.

El presente trabajo muestra el modelo completo del experimento original y analiza la estrategia de los agentes para resolver dicha tarea. Se presume que el hecho de que exista un mayor número de clicks en los agentes, comparado con los demás objetos del ambiente, revela que existe un auténtico reconocimiento de explícito de agencialidad. Un modelo con el reconocimiento explícito entre agentes serviría como otra forma de comprobar el rol que juegan las interacciones en la identificación de otro agente. Además de reconocer la importancia de la generación de modelos sintéticos para el estudio de comportamientos cognitivos, ya que éstos están basados en sistemas dinámicos, los cuales se comportan de manera muy similar a los eventos que suceden en la naturaleza.

1.1. Objetivo general

Implementar un modelo de robótica evolutiva que demuestre la detección explícita de agencialidad por medio de clicks.

1.1.1. Objetivos específicos

- Reproducir los modelos anteriores de robótica evolutiva inspirados por el experimento del cruce perceptual.
- Identificar y entender los diferentes procedimientos utilizados para la generación de modelos mínimos de cognición.
- Interpretar y analizar el comportamiento dinámico de los sistemas mínimos de cognición.
- Incluir el comportamiento del click en el modelo de robótica evolutiva.
- Demostrar que el comportamiento del click proporciona mayor información sobre el reconocimiento explícito de agencialidad.
- Sustentar que la robótica evolutiva es una herramienta importante para el análisis de la cognición social. Verificar que el estudio de la cognición social se ve beneficiado de herramientas como la teoría de sistemas dinámicos. Lo que de esta forma permitiría que se replantee el enfoque con el que se aborda el estudio de fenómenos de cognición social.

1.2. Metodología

La robótica evolutiva fue elegida como la metodología adecuada para desarrollar un modelo mínimo de cognición social. Esta metodología comenzó a utilizarse a inicios

1. INTRODUCCIÓN

de 1990 arrojando resultados muy satisfactorios y prometedores para el estudio de la cognición. Empleada inicialmente para la optimización de controladores para robots físicos (Nolfi & Floreano, 2000), posteriormente comenzó a aplicarse en el área de la ciencia cognitiva para el progreso en cuestiones teóricas y conceptuales (e.g Di Paolo, 2000; Iizuka & Ikegami, 2004, 2007; Iizuka & Di Paolo, 2007; Di Paolo et al., 2008; Froese & Di Paolo, 2008). En el capítulo 4 se habla con más detalle sobre las características de esta metodología y su correspondiente implementación en este trabajo.

1.3. Contribuciones

Los modelos anteriores de comportamiento mínimo cognitivo demuestran que el papel de la interacción perceptual conjunta tiene un rol fundamental para reconocer a otra entidad con intenciones propias. Así mismo, para lograr un mejor entendimiento del fenómeno de contingencia social no basta con analizar solamente las capacidades individuales de los agentes, sino también las propiedades de interacción conjunta de todo el sistema. El análisis dinámico de esos modelos demostró que las estrategias de los agentes pueden ser mejor entendidas desde una perspectiva dinámica. Ahora lo que corresponde estudiar es cómo se modifican esas estrategias cuando existe un verdadero reconocimiento explícito entre los agentes. Además, el modelo puede ayudar a tomar una perspectiva distinta de análisis en las áreas de neurociencias y psicología principalmente. Aparte contribuye a generar nuevos experimentos para el estudio de cognición social y ampliar la aplicación de los sistemas complejos para conocer más sobre los mecanismos subyacentes a dichos fenómenos.

1.4. Estructura de la tesis

En el capítulo 2 se abordan los orígenes y las diferentes perspectivas para el estudio de la mente. Desde la filosofía con el enfoque dualista de Descartes, pasando por la psicología y el conductismo como la corriente más fuerte que tuvo en sus inicios. Después se habla sobre el surgimiento de la cibernética enfocada en dar una solución a la creación de robots y -en términos más generales- de sistemas lo suficientemente robustos y flexibles ante las contingencias a las que se enfrentan en la realidad. Posteriormente, cuestionamientos sobre los desafíos que enfrentaba la cibernética terminarían dirigiéndose y concentrando sus recursos en el desarrollo de sistemas basados fuertemente en procesos computacionales como los descritos por Turing. Eventualmente, surgieron complicaciones para alcanzar objetivos mucho más ambiciosos y eso generó fuertes críticas hacia esta perspectiva.

Con el fin de poder generar un paradigma con la capacidad de dar soporte a los retos planteados en el Capítulo 3 se explican cómo surgieron nuevos marcos de referencia. En punto la inteligencia dio un giro importante al integrar características como la estructura física de los robots y también el espacio físico en el cual es necesario que

ejecuten determinadas tareas. En esta misma línea hubo un desarrollo mucho más completo con el enactivismo, paradigma que retoma los trabajos de la inteligencia artificial corporizada, situada y embebida. Las investigaciones basadas en este enfoque muestran la importancia de establecer características imprescindibles retomadas de los organismos vivos que no habían sido contempladas en los enfoques anteriores.

En el Capítulo 4 se presentan las metodologías y herramientas utilizadas para el estudio y análisis del modelo generado. Primero se habla sobre la robótica evolutiva, su aplicación en la creación y diseño de robots. Más adelante se explican dos elementos fundamentales de esta metodología: las redes neuronales recurrentes de tiempo continuo y los algoritmos genéticos. Otra herramienta que se menciona en este capítulo es la teoría de sistemas dinámicos empleada para entender los mecanismos que subyacen al comportamiento de sistemas con interacciones. Se parte de la descripción de conceptos básicos, después se menciona el enfoque de los sistemas dinámicos de manera general y se finaliza con la adopción de esta perspectiva en los sistemas cognitivos.

Las herramientas mencionadas en el párrafo anterior comenzaron a tener bastante aplicación en el área de las ciencias cognitivas, en lo cual se enfoca el Capítulo 5. Al comienzo se da una introducción a la cognición social. La cognición social es un fenómeno integrado por una gran cantidad de elementos, debido a ésto se crearon modelos donde se puede observar cognición social sin que sean sumamente complicados y difíciles de implementar. A razón de lo anterior, surge el paradigma de cruce perceptual, el cual busca estudiar los mecanismos que permiten que una persona pueda reconocer a otra -e incluso interactúe con ella- de entre los demás objetos presentes. También se habla de modelos artificiales que buscan simular los modelos anteriores hechos con personas. El capítulo concluye con un recorrido por los modelos sintéticos inspirados en el paradigma del cruce perceptual.

El objetivo del Capítulo 6 es describir con detalle la parte principal de esta tesis: un modelo de simulación donde se despliega un reconocimiento explícito de agencialidad. Primero se habla sobre el diseño del modelo, el ambiente virtual, las características de los objetos y la estructura física de los agentes, etc. Después viene la parte del procedimiento donde se plantea la función de aptitud diseñada para los agentes artificiales, así como también especificaciones del algoritmo genético utilizado para obtener los parámetros óptimos para los controladores de los agentes. También se utiliza la teoría de los sistemas dinámicos para analizar en conjunto la dinámica de los controladores de los agentes cuando se encuentran en el proceso de interacción.

Finalmente el Capítulo 7 contiene las conclusiones generales de esta tesis, estableciendo que es importante dejar de abordar el aspecto cognitivo de manera aislada. Hay aspectos que no pueden ser descubiertos o identificados estudiando a los individuos de manera personal, sino que es imprescindible que esos individuos se encuentren interactuando y que su análisis se haga durante esa interacción. Los nuevos marcos de referencia que fueron utilizados en esta tesis han sido constituidos como tales en las últimas décadas y aún hace falta que se haga más investigación para fortalecer sus bases.

La ciencia cognitiva

2.1. Orígenes

La inquietud de poder entender y también explicar los fenómenos que percibe el ser humano ha sido una constante a lo largo de la historia de la humanidad. Uno de esos fenómenos que ha sido de particular interés es el proceso de pensamiento, que a su vez abarca otros aspectos como: el cerebro, la mente, el conocimiento y la capacidad de razonar, sólo por mencionar algunos. Distintas áreas han surgido con el fin de estudiar de manera formal y sistemática estos fenómenos. La siguiente sección presentará una breve perspectiva de lo que es la ciencia cognitiva, incluyendo un breve resumen de las disciplinas que más la han influenciado. En las secciones posteriores se irán analizando diferentes eventos que impactaron de manera relevante al desarrollo de esta ciencia y que son de gran importancia para el trabajo desarrollado en esta tesis.

2.1.1. Definición y disciplinas subyacentes

La ciencia cognitiva es el estudio interdisciplinario de la mente y el cerebro, desarrollada sobre los pilares de la filosofía, psicología, inteligencia artificial, neurociencia, lingüística y antropología (Fig. 2.1).

Desde la perspectiva filosófica, la ciencia cognitiva se ha visto beneficiada como resultado de los cuestionamientos profundos acerca de las implicaciones que tienen los distintos enfoques para el estudio de la mente. La filosofía se ha ocupado de preguntas que tienen que ver con la relación entre la mente y el cuerpo, y con cuestiones metodológicas acerca de la naturaleza de la mente, sus diferentes estados, su funcionamiento, sus propiedades y la conciencia. Una de las aportaciones de la filosofía a la ciencia cognitiva es la generalidad. Por generalidad se hace referencia a que las reflexiones y análisis filosóficos intentan responder preguntas que son más amplias que las que se abordan en otras disciplinas más especializadas como la psicología, la neurociencia, la lingüística, la antropología y la inteligencia artificial. La generalidad filosófica es crucial en un campo interdisciplinario como la ciencia cognitiva, ya que aborda preguntas

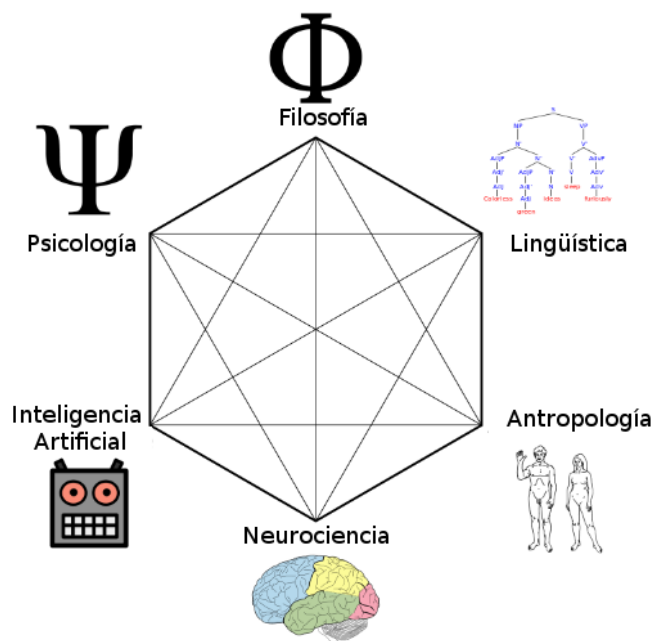


Figura 2.1: Disciplinas que contribuyeron al nacimiento de la ciencia cognitiva. Traducción de la imagen hecha por Lowe & Rotenberg (2012)

que combinan diferentes líneas de investigación al mismo tiempo. Lo anterior ayuda a unificar lo que desde otra perspectiva podrían percibirse como enfoques diferentes para entender la mente.

Por otro lado, la psicología definida en 1980 por William James como "la ciencia de la mente, abarcando tanto sus fenómenos como sus condiciones" (Russell, 1987/2004) tiene como principal enfoque el aspecto práctico del estudio de la mente, utilizando a la experimentación con seres humanos como método primordial. En términos muy generales la psicología se ha especializado en efectuar experimentos que se llevan en laboratorios bajo condiciones controladas. Básicamente y de manera general en esos experimentos se realizan pruebas comparativas entre sujetos acerca de su desempeño en cuestiones como: la respuesta de los sujetos frente a determinados estímulos como imágenes, su habilidad para tener un razonamiento lógico, acondicionamiento, etc. Los experimentos psicológicos que abordan de manera sistemática el funcionamiento mental desde diversas direcciones son cruciales para que la ciencia cognitiva cumpla la característica de ser una ciencia.

Muy parecido al enfoque de la psicología lo encontramos en el área de neurociencias, donde también se realizan experimentos en laboratorios. Sin embargo, el enfoque de estos experimentos es distinto. Aquí se busca un análisis más orientado a la parte biológica del cerebro: neuronas, sinapsis, estructuras neurales, patrones de activación.

Para eso, se realizan estudios empleando diferentes técnicas: imagen por resonancia magnética funcional, tomografía por emisión de positrones, tomografías computarizadas, electroencefalografía (Marter & Shieh, 2010). Los análisis con las técnicas anteriormente mencionadas permiten identificar distintas zonas de activación al ejecutar tareas específicas. De esta forma, se han podido identificar áreas en el cerebro especializadas en ciertas tareas, por ejemplo: la zona del lenguaje, la memoria, la visión, la actividad motriz, etc. Además se han generado desarrollos de tratamientos para personas que han sufrido daños cerebrales o que sufren alguna otra patología.

En el área de la antropología se ha analizado la mente desde el resultado del pensamiento humano y cómo se desarrolla desde la parte cultural. Desde esta perspectiva, se toma en consideración un aspecto que en otras disciplinas relacionadas no se le da tanta ponderación: el aspecto social. La antropología cognitiva se refiere al “estudio de la relación entre la sociedad y el pensamiento humano” (D’Andrade, 2003). Se busca el tratar de comprender cómo las personas que provienen de diferentes sociedades hacen uso de su conocimiento en la vida cotidiana. Así mismo, se trata de determinar de qué forma las comunidades organizan su conocimiento, cómo lo transmiten y lo mantienen, además de saber cómo lo unifican para que eventualmente éste forme parte de la cultura misma.

Otra disciplina que se ha hecho presente en la ciencia cognitiva es la lingüística. La lingüística es la disciplina científica interesada en el estudio del lenguaje humano y tiene sus bases más formales en los trabajos desarrollados por Noam Chomsky a mediados del siglo XX. Una cuestión que ha sido de particular interés para los lingüistas es la relación que guarda el lenguaje con la cognición, teniendo como resultado algunas preguntas como: ¿las personas que hablan diferentes idiomas piensan diferente?, ¿es fundamental un cierto nivel de desarrollo cognitivo para aprender un idioma?, ¿el lenguaje es una capacidad mental diferente a otras habilidades cognitivas o es fundamentalmente diferente? En la búsqueda de las respuestas a estas preguntas la lingüística se ha visto relacionada en gran medida con la ciencia cognitiva proponiendo sus propios marcos de referencia.

2.1.2. Dualismo y conductismo

Los primeros estudios formales sobre la mente nos remontan a los grandes filósofos de la historia como Platón y Aristóteles, los cuales comenzaron a explicar la naturaleza del conocimiento humano. Las ideas más relacionadas al aprendizaje y al conocimiento en el ser humano las podemos encontrar en el diálogo socrático *Meno* escrito por Platón. Es este texto Platón propone la idea de que los seres humanos poseen el conocimiento de encarnaciones pasadas y que el aprendizaje consiste en redescubrir ese conocimiento dentro de nosotros. Por parte de Aristóteles en su texto filosófico *De Anima* habla sobre la naturaleza de las cosas vivas. Según el planteamiento de Aristóteles las plantas poseen la capacidad para nutrir y reproducirse, los animales adicionalmente tienen sentido de percepción y acción, y finalmente los seres humanos tienen todas las características anteriores además de intelecto. Posteriormente adquirió gran fuerza el

2. LA CIENCIA COGNITIVA

dualismo cartesiano durante el siglo XVIII, postura que básicamente sostiene que la mente y el cuerpo son, en cierto sentido, radicalmente dos cosas distintas. El dualismo está estrechamente relacionado con el pensamiento de René Descartes quien sostiene que existen dos tipos de sustancias: la materia, que tiene la propiedad esencial de ocupar un espacio; y la mente, que posee la propiedad de pensar (Robinson, 2016). De esta forma los cuestionamientos y el estudio de la mente quedaron, al inicio, principalmente en el ámbito de la filosofía.

No fue sino a partir del siglo XIX con el surgimiento de la psicología experimental que se comienza a generar un enfoque distinto y mucho más sistemático. Un evento muy importante fue la inauguración del primer laboratorio experimental en psicología en la Universidad de Leipzig, Alemania, laboratorio dirigido por el que después sería considerado como el padre de la psicología Wilhem Wundt (Fraisie et al., 1963). Durante varias décadas la psicología experimental estuvo dominada por el conductismo, un enfoque sistemático enfocado en el comportamiento de organismos individuales. El conductismo está basado en la postura de que el comportamiento son actos reflejos producidos como respuesta a un estímulo del ambiente. Esta idea tiene dos puntos muy importantes: primero, niega que la psicología sea una ciencia que estudie la mente; segundo, dado que el comportamiento está determinado por cuestiones externas (ambiente), entonces no es necesario analizar los diferentes estados de la mente o los procesos psicológicos internos del individuo (Graham, 2016).

Entre los años 1920s y 1950s la psicología en América estuvo dominada por el conductismo. Sin embargo, después de esta época se suscitaron dos descubrimientos que serían de importante relevancia en el área de la psicología: primero, Paul Broca descubre el área del cerebro responsable de la producción de lenguaje y segundo (Malone, 2009), Carl Wernicke descubre el área que se cree es la principal responsable de la comprensión del lenguaje (Eysenck, 1990). Durante la Segunda Guerra Mundial hubo un creciente interés para obtener un mayor entendimiento acerca del desempeño en los seres humanos. Cuestiones como encontrar la mejor manera de entrenar a los soldados para utilizar nueva tecnología, así como lidiar con asuntos de atención bajo coacción se convirtieron en áreas de desarrollo en el ámbito militar. En este sentido, el conductismo dejó de hacer grandes aportaciones y eventualmente comenzó a adquirir fuerza el enfoque de teoría de la información para explicar el comportamiento y pensamiento humanos, del cual se hablará posteriormente en la siguiente sección.

Como ya fue mencionado previamente, la filosofía y la psicología no serían las únicas áreas que interesadas en aspectos relacionados con la mente, el cuerpo y su relación con el entorno. Posteriormente se gestaría una nueva área conocida como cibernética surgiendo de la necesidad de nuevas herramientas para abordar la complejidad, interdependencia y la adaptación de sistemas. En la siguiente sección se presenta una breve historia de la cibernética así como de sus aportaciones y el camino hacia la ciencia cognitiva.

2.2. La cibernética

El término “cibernética” fue introducido en el siglo XX por Norbert Wiener en su libro *Cybernetics: Or control and communication in the Animal and the Machine*. De acuerdo con Wiener (1948), la cibernética nació como resultado de tres líneas de investigación:

1. Teoría de la Información.
2. Redes Neuronales.
3. Conexión entre retroalimentación negativa y comportamiento intencional.

El enfoque de la teoría de la información que fue adoptado por Wiener estuvo inclinado hacia la ingeniería, postura bastante similar a la desarrollada por su colega Claude Shannon considerado “el padre de la teoría de la información” (James, 2009). Shannon propuso que la información era comunicada a través del envío de señales mediante una serie de etapas de transformaciones. Lo anterior sugería que, tanto la percepción y la memoria en los seres humanos podrían estar conceptualizadas de una manera similar: la información sensorial entra a los receptores, luego pasa por los analizadores perceptivos, cuyas salidas en turno son la entrada de los sistemas de memoria. Estos puntos serían la base del enfoque de “teoría de la información”: la cognición puede ser entendida como un flujo de información que ocurre dentro del organismo.

Por otro lado la cuestión de las redes neuronales está relacionada con el trabajo de Warren McCulloch y Walter Pitts (1943): “Un cálculo lógico de las ideas inmanentes en la actividad nerviosa”.

El trabajo realizado por Wiener en la cibernética ayudó a formalizar conceptos relacionados al control de retroalimentación. Profundizando más en este aspecto, el significado básico de “retroalimentación” es simplemente: algo que es producido por una máquina o un organismo y que es dirigido de regreso para modificar el proceso de producción. Si incrementa la salida del proceso, entonces es llamado “retroalimentación positiva”. Si la retroalimentación es utilizada para regular o limitar el proceso que lo genera, entonces es una retroalimentación negativa”. Este segundo tipo de retroalimentación constituye el punto central de los mecanismos de control que fueron de gran interés en las primeras fases de la cibernética.

Un ejemplo muy representativo de retroalimentación negativa es el regulador de Watt, cuyo propósito es mantener la velocidad de la rueda de un motor de vapor a un valor dado a pesar de las perturbaciones. El regulador de Watt, consistía en un conjunto de pesos giratorios impulsado hacia el exterior por la fuerza centrífuga que “sensa” la velocidad del motor. El “sensado” es, por supuesto, puramente metafórico. El regulador como tal no posee órganos de los sentidos, pero está construido de tal manera que, al llegar a una cierta posición, puede cerrar o abrir una válvula por medio de una conexión física de palancas o cadenas.

En todos estos aparatos, la retroalimentación es mecánica y no implica señales o comunicación simbólica. Otro aparato producto de los trabajos de la cibernética fue el termostato, el cual fue utilizado por los primeros cibernéticos como el ejemplo más representativo del área, ya que representaba de manera práctica conceptos teóricos fundamentales para la cibernética. Ejemplificando el funcionamiento del termostato en un sistema de aire acondicionado su está para mantener la temperatura en un espacio cerrado en el nivel deseado. Es decir, un individuo establece una temperatura específica como valor de referencia, y el termostato “detecta” la temperatura actual por medio de un termómetro y tiene la capacidad para compararlo con el valor establecido, si lo que se registra es inferior al de referencia, se activa el calentador, si es más alto, se activa el sistema de refrigeración. Los primeros investigadores en el área de la cibernética generalizaron posteriormente la idea de la retroalimentación negativa para abarcar cualquier comportamiento que tuviera la naturaleza de buscar un propósito o una meta. Lo que ilustran muy bien Rosenblueth y sus colegas en *Behavior, purpose and teleology* (Rosenblueth et al., 1943): “Todo comportamiento que posee un propósito se podría considerar que requiere una retroalimentación negativa. Si se requiere alcanzar una meta, algunas señales de la meta son requeridas en algún momento para dirigir el comportamiento.”. Después de los avances mencionados vendría el trabajo revolucionario de R. Ashby para marcar un parteaguas en la cibernética, poniendo a prueba la representación explícita del objetivo que el sistema busca alcanzar y que era una constante en todos los proyectos. En la siguiente sección se habla sobre las aportaciones de Ashby que marcarían el inicio de una nueva era.

2.2.1. Cibernética de Ashby

Se daría un gran giro a la cibernética con los avances realizados por William Ross Ashby, con su obra publicada en el año 1952 con el título *Design for a brain* (1952), en donde Ashby describió su célebre homeostato: dispositivo electrónico autorregulado por retroalimentación. En ese libro también expuso su novedosa concepción de un cerebro artificial, contribuyendo de forma esencial a la consolidación de la cibernética moderna como ciencia. Más tarde con su obra *An Introduction to Cybernetics* (Ashby, 1956) estableció los principios generales para la creación de mecanismos de funcionamiento análogos al cerebro humano que sirvieran de ayuda a la inteligencia. Adicionalmente ese libro trataba sobre los procesos homeostáticos presentes dentro de organismos vivos en vez de en ambientes electrónicos o de ingeniería.

Retomando el homeostato de Ashby, este fue creado como una prueba de concepto para probar que lo que puede parecer un comportamiento inteligente para un observador externo puede ser explicado utilizando un sistema matemático simple impulsado por funciones de paso aleatorias. Cuando el homeostato fue presentado ante la comunidad científica, no hubo gran aceptación del mismo, ya que éste no contaba con una representación explícita sobre el objetivo o propósito del sistema. Sin embargo, el dispositivo contaba con la capacidad de adaptarse por sí mismo al ambiente. Además de exhibir un comportamiento de aprendizaje a través de su capacidad para mantener la

homeostasis en un entorno cambiante. Lo anterior implicó se que cuestionara fuertemente esa característica fundamental en los trabajos anteriores: el representacionalismo explícito (Froese, 2010).

Este proyecto generó un cambio radical en la cibernética dado que se generaron dos retos muy importantes: la auto-organización y el rol del observador. En la búsqueda de dar solución a esos retos surgieron conceptos muy relevantes como el de la autorreferencia (que más tarde sería transformado en autopoiesis y cerradura organización) y la auto-organización.

2.2.2. Cibernética de segundo-orden

La relación de la cibernética de Ashby con la cibernética de segundo orden es como la relación entre la visión newtoniana del universo y la visión propuesta por Einstein. Justo como la descripción de Newton permanece totalmente apropiada y aplicable en diferentes escenarios (incluyendo viajes a la luna), la cibernética de primer orden mantiene su valor y frecuentemente nos proporciona las herramientas necesarias para resolver ciertos problemas (por ejemplo, el desarrollo de dispositivos de control). Y justamente como la visión newtoniana es entendida como una versión particular, simplificada y restringida de la visión de Einstein, la cibernética de primer orden es una versión particular, simplificada, restringida (y lineal) de la cibernética de segundo orden. Aunque la analogía va incluso más allá. Una diferencia en ambos casos, en la visión de Einstein y en la cibernética de segundo orden, es que el observador -esencialmente excluido de la visión newtoniana y de la cibernética de primer orden- es incluido. La cibernética de segundo orden presenta un (nuevo) paradigma en el cual la relación que tiene el observador con lo observado es mucho mejor entendida de una manera circular, es decir, el observador deja de ser algo neutral y desprendido de su entorno y lo que es analizado/entendido es el sistema completo de observación.

Un personaje importante para el desarrollo de la cibernética de segundo orden fue Stafford Beer quien se enfocó en la organización de grandes sistemas de gran complejidad. También desarrolló su trabajo basado en la “Ley de Variedad Requerida”: para que cualquier sistema pueda controlar efectivamente otro sistema (no restringiendo a priori sus posibles salidas) el primero deberá tener al menos tanta variación como el segundo, considerando variación como una medida del número de los posibles estados que un sistema puede alcanzar (Ashby, 1956, p. 207).

La cibernética de segundo orden le da un papel relevante al observador. Una consecuencia inmediata de lo anterior es que todo el conocimiento es visto como algo dependiente del involucramiento del observador. Preguntas como las siguientes surgen como consecuencia de darle mayor importancia al observador: ¿existe un objeto en una realidad externa? de ser así, ¿qué podemos saber de eso, dado que nuestro conocimiento depende de nosotros y nunca podremos prescindir de nuestra presencia? Este no es un problema nuevo, enfoques anteriores tienen la tendencia de, ya sea negar el problema y establecer que “la realidad allá afuera” es obvia, o desde el solipsismo, negar cualquier cosa fuera de la sensación y entendimiento del individuo.

Por otro lado, otro expositor muy importante de la cibernética de segundo orden fue Heinz von Foerster (1979) quien publicó *Cybernetics of cybernetics* donde propuso un lúcido argumento acerca de complejidad, modelación, retraso y sobrecarga de información.

Las contribuciones más relevantes elaboradas por Von Foerster fueron la introducción de tres conceptos importantes:

- Empleo de la cibernética para aplicar sus avances a la cibernética misma. Es decir, la cibernética aplicada a así misma en un segundo orden de operación.
- Integración del observador.
- Cálculo de realidades estables.
- Y por último, es que nosotros construimos nuestras realidades.

Ahondando en los últimos dos puntos, von Foerster trabajó en sistemas auto-organizantes con el fin de saber cómo se podrían calcular realidades estables (lo que él llamaría “objetos”) a través de la observación recursiva. En esa misma línea también encontró un cierto tipo de funciones que, aplicadas recursivamente llegan a estados estables y (dinámicamente) auto-perpetuables. Además von Foerster tenía especial interés en la relación entre estos mecanismos y el funcionamiento del sistema nervioso.

Retomando el trabajo de Ashby, la creación del homeostato y en general sus aportaciones implicaron un importante reto para la cibernética. Esta situación propició el surgimiento de un enfoque basado en la información que finalmente se transformaría en la inteligencia artificial simbólica sobre lo cual se hablará en la siguiente sección.

2.3. Inteligencia artificial simbólica

2.3.1. Teoría computacional de la mente

Alrededor de los años 1960s y 1970s tuvo mucho auge la idea de que la mente por sí misma podía ser entendida como un sistema computacional. Partamos de una cuestión esencial y básica para la computación: el algoritmo. En términos generales, un algoritmo es un procedimiento explícito constituido por una serie de pasos con el objetivo de resolver una problema. El resultado es una rutina mecánica de instrucciones que determina cómo proceder en cada paso. Seguir esas instrucciones no requiere de ingenio o creatividad para llevarse a cabo, lo único que es necesario es que sean ejecutadas de la forma indicada. Hasta inicios del siglo XX, los matemáticos se basaron en nociones informales de algoritmos y computación sin necesidad de llegar a crear análisis sistemáticos más formales. El desarrollo en los fundamentos de las matemáticas eventualmente llevó a los matemáticos a la búsqueda de procedimientos mucho más sistemáticos.

Uno de los intentos que hasta ahora permanece muy presente y que ha tenido un enorme impacto para las bases de la computación fue *la máquina de Turing* propuesta

por Alan Turing en su artículo: *On Computable Numbers, with an Application to the Entscheidungsproblem* (Turing, 1936). La máquina de Turing es un modelo abstracto de un dispositivo computacional idealizado en tiempo y espacio de almacenamiento. El dispositivo manipula símbolos que son tomados de un alfabeto finito y que pueden ser escritos o borrados sobre las localidades en una memoria infinita. Turing argumentó que cualquier algoritmo simbólico ejecutado por un humano puede ser replicado por una máquina de Turing adecuada. Además, concluye de que a pesar de que su propuesta no es compleja, sí es lo suficientemente poderosa para capturar ciertos procedimientos mecánicos que pueden ser ejecutados por seres humanos a través de configuraciones simbólicas. Finalmente el trabajo de Turing permitió crear las bases de la ciencia de la computación, la cual busca construir, estudiar y entender los sistemas computacionales.

2.3.2. Representación del conocimiento y computacionalismo

A partir de lo desarrollado por Turing, comenzaron a surgir preguntas como: ¿es posible crear una máquina que pueda pensar? ¿la mente por sí misma es una máquina pensante? ¿somos capaces de poder crear una computadora con la habilidad de pensar?

Con todas estas ideas en mente da surgimiento la Inteligencia Artificial, disciplina que se establece formalmente en la conferencia Dartmouth en 1956 organizada por Marvin Minsky, John McCarthy, Claude Shannon y Nathan Rochester y donde se considera que fue utilizado por primera vez el término de Inteligencia Artificial (McCarthy et al., 1955). En dicha conferencia se popularizó la metáfora de la computadora digital para referirse a la mente humana, propuesta sería abordada desde distintos ámbitos. Por ejemplo, Allen Newell y Herbert A. Simon quienes entre 1955 y 1956 desarrollaron un programa de computadora llamado “Teórico de la Lógica”, el cual fue el primer programa que imitaba la habilidad de resolución de problemas. De hecho, terminaría resolviendo 38 de los primeros 52 teoremas del *Principia Mathematica* e incluso encontraría nuevas y elegantes maneras para llegar a la solución de algunos de ellos (McCorduck, 2004). Posteriormente en 1959, desarrollarían el Solucionador General de Problemas cuyo objetivo era convertirse en una máquina universal solucionadora de problemas (Newell et al., 1959).

Sus trabajos buscaron la forma de explicar cómo es que se da la resolución de problemas en el ser humano: cuáles son los procesos utilizados y qué mecanismos desempeñan dichos procesos (Newell & Simon, 1971). Fodor, por otro lado enfoca su atención a la manipulación de símbolos durante el proceso computacional de una máquina de Turing. Fodor postula un sistema de representación mental incluyendo tanto representaciones primitivas como complejas. Además, plantea el fenómeno mental como la combinación de toma de decisiones, percepción y procesamiento lingüístico (Fodor, 2000).

Los enfoques teóricos relacionados a la representación del conocimiento y el computacionalismo son los siguientes:

- Lógica formal: Es la rama de la lógica enfocada exclusivamente en los principios del razonamiento deductivo. Las proposiciones y los predicados utilizados por la

lógica formal la convierten en una herramienta capaz de representar conocimiento complejo.

- Reglas: Son guías establecidas para determinar una conducta o una serie de acciones. En el ámbito computacional existen sistemas basados en reglas, los cuales han permitido mejorar procesos de aprendizaje y desarrollar un subconjunto de los llamados sistemas inteligentes.
- Conceptos: Es una idea abstracta que representa las características fundamentales de aquello que nombra y son un tipo muy importante de representación mental.
- Analogías: Es un proceso cognitivo de transferencia de información o significado de un tema en particular (analogía o fuente) hacia otro (el objetivo).
- Imágenes: Son artefactos que representa la percepción visual, por ejemplo, una imagen bidimensional que tiene una apariencia similar a un sujeto -por lo general un objeto o una persona- proporcionando así una representación de la misma.
- Modelo Bayesiano: Es un enfoque que está relacionado con el análisis racional de la cognición a través del uso de la inferencia bayesiana y el modelado cognitivo. Se entiende por inferencia bayesiana al método de inferencia estadística en el cual se utiliza el teorema de Bayes para actualizar la probabilidad de una hipótesis a medida de que se disponga de más evidencia o información. El trabajo del enfoque bayesiano consiste en probar la hipótesis de que los sistemas cognitivos se comportan como agentes racionales bayesianos en ciertos tipos de tareas (categorización, lenguaje, control motor, aprendizaje secuencial, aprendizaje por reforzamiento, teoría de la mente).

2.3.3. Argumentos y críticas sobre el representacionalismo y la teoría computacional de la mente

La teoría computacional de la mente determina que el funcionamiento de la mente está basado en procesos como la lógica de proposiciones, reglas, conceptos, razonamiento matemático, etc. Este paradigma permitió que el estudio de la mente fuera más allá del enfoque conductista -como fue explicado en la sección previa- y tomara un formalismo científico para el estudio de la cognición. Además ha tenido un gran éxito en el desarrollo de una amplia variedad de aplicaciones en el ámbito de la computación: desde el desarrollo de sistemas de cálculos numéricos pasando por los sistemas organizacionales, además de tener un gran impacto tanto en las ciencias exactas como en las ciencias sociales, y un sin fin más de ejemplos que utilizamos de manera cotidiana. Sin duda alguna muchas disciplinas se han visto beneficiadas de esta poderosa herramienta que ha logrado poner al alcance capacidades de cómputo que han permitido entender fenómenos que hace años permanecían alejados de nuestro entendimiento.

A pesar de todas la aportaciones mencionadas, esta metáfora de la mente como un sistema computacional ha tenido fuertes críticas desde distintas perspectivas, principalmente por las limitaciones que se han encontrado para poder describir cuestiones relacionados a la cognición. A continuación se listan algunas de las implicaciones y cuestionamientos que han surgido al haber tenido como base este marco de referencia (Thagard, 2014).

- Las máquinas de Turing ejecutan símbolos, es decir las entradas y las salidas son símbolos contenidos en localidades de memoria. A diferencia de la mente que recibe estímulos sensitivos como entradas y produce salidas motoras.
- La máquina de Turing tiene una memoria con capacidad infinita, mientras que los organismos biológicos poseen una capacidad limitada.
- La máquina de Turing tiene un “procesador” central que opera de manera serial, ejecutando una instrucción a la vez. Mientras que los procesos naturales son típicamente desordenados y tienen un funcionamiento en paralelo.
- La manera en que la máquina de Turing determina el siguiente paso sobre la base lógica es por medio de un algoritmo basado en reglas.
- El problema del marco en inteligencia artificial: ¿Es posible, en principio, limitar el alcance de razonamiento requerido para deducir las consecuencias de una acción? (McCarthy & Hayes, 1969).
- El problema del significado de los símbolos: ¿Cómo llega un sistema inteligente a entender los conceptos que utiliza? (Harnad, 1990).
- ¿Hasta qué punto el conocimiento viene de manera innata o es adquirido por la experiencia?
- ¿El entendimiento que tiene una persona acerca de otra consiste en tener una teoría de la mente o simplemente es capaz de simular a la otra persona?
- *Imágenes de la mente*: ¿La mente piensa a través de imágenes o solamente a través del lenguaje utilizando representaciones?
- *Significado*: ¿Cómo adquieren significado las representaciones mentales? ¿Hasta qué punto el significado de una representación depende de su relación con otras representaciones y con el mundo?
- *Identidad mente-cerebro*: ¿Los estados mentales son también estados del cerebro? ¿O pueden ser múltiplemente realizados por otros estados mentales? ¿Cuál es la relación entre psicología y neurociencia?
- *El libre albedrío*: ¿Las acciones humanas son libres o son una consecuencia de procesos cerebrales bien definidos?

2. LA CIENCIA COGNITIVA

- *Psicología moral*: ¿Cómo es que la mente/cerebro realiza juicios éticos?
- *El significado de la vida*: ¿Cómo pueden las mentes, construidas de manera biológica como cerebros, encontrar un valor y un significado?
- *Las emociones*: ¿Qué son las emociones y qué rol juegan en el pensamiento?
- *Enfermedades mentales*: ¿Qué son las enfermedades mentales y cómo son relevantes los procesos psicológicos y neurológicos en su explicación y tratamiento?
- *La apariencia y la realidad*: ¿Cómo es que la mente/cerebro genera y evalúa representaciones del mundo exterior?
- *Ciencias sociales*: ¿Cómo la explicación de la operación de la mente interactúa con la explicación de la operación de grupos y sociedades?

Con el auge que tuvo la inteligencia artificial simbólica, como ya fue descrito, la mente se vio reducida a procesos como la capacidad de razonamiento, utilización de un lenguaje complejo, habilidad para generar deducciones, cálculos matemáticos y en general se tomaron las cuestiones complejas-abstractas como las más generales. Lo anterior eventualmente llevó a generar sistemas computacionales que imitaran de una manera muy precisa esas habilidades. Sin embargo, para entender la mente, este marco de referencia deja fuera muchos aspectos del ser humano, e incluso si tomáramos una perspectiva más amplia, de un ser vivo o un agente. Un agente es caracterizado porque realiza acciones guiadas por metas o fines, además de que un agente se ve inmerso en un mundo y la relación que va generando con él. Entonces, para el agente, sus acciones, las consecuencias de ellas y su relación con el entorno implican un significado, aspectos que para una computadora o un sistema computacional no tienen sentido, ya que la computadora sólo ejecuta una serie de instrucciones bien definidas y no le importa si ha terminado bien, o el mismo significado de los resultados de su ejecución; más bien quién le genera un significado a todo eso es el usuario de la computadora.

Otro aspecto que ha tratado de ser imitado por la Inteligencia Artificial es el tema del aprendizaje. Hasta ahora se han creado grandes sistemas que, por ejemplo, han aprendido a jugar ajedrez y *go*, derrotando a los seres humanos más hábiles para esos juegos sobre la faz de la tierra. Sin embargo, las características de ese tipo de aprendizaje son bastante determinadas, es decir, el aprendizaje se da bajo condiciones y contextos bien definidos. Además de que ese aprendizaje se reduce a un aprendizaje de patrones, de formas, basado en reglas de lógica, y aún cuando el espacio de búsqueda es inmenso como en los juegos de ajedrez y *go*, dista mucho de ser como el mundo abierto en el que vivimos.

Dada la limitación de la propuesta de la inteligencia artificial para estudiar la cognición cobraron fuerza otros enfoques que tomaron en cuenta más aspectos para el estudio de la mente. Además, comenzaron a gestarse cuestionamientos relacionados con las herramientas que el enfoque computacional ofrece. Algunos de los desafíos que se presentaron para la inteligencia artificial simbólica -y que aún siguen vigentes- fueron: ¿qué tan robusto puede ser un sistema para mantenerse estable en situaciones

de ruido y perturbaciones?, ¿qué sucedería si ocurre un evento que no haya sido contemplado durante el diseño del sistema?. Generalmente los sistemas desarrollados por la inteligencia artificial simbólica terminan abruptamente su funcionamiento debido a la falta de instrucciones que indiquen qué acción realizar en caso de una contingencia. Otro desafío para este tipo de sistemas es su desempeño en tiempo real y las variaciones que se presentan debido a que el mundo exterior presenta variaciones constantemente. Lo que va muy relacionado con el tema del aprendizaje y adaptación al cambio, ya que los mecanismos de estos sistemas son específicamente diseñados por el programador del sistema y bajo condiciones específicas.

Lo anterior plantea un gran reto frente a acciones que los seres humanos realizan de manera cotidiana, como el hecho de que un niño pueda explorar y percibir el mundo sin hacer un gran esfuerzo, reconocer las emociones de las personas que se encuentran a su alrededor, etc. Hasta el momento éstas y otras características del ser humano no ha podido lograrse con una computadora. Estos desafíos sugieren que la intuición, creatividad o diversas habilidades del ser humano parecen resistirse al formalismo de un sistema computacional (Dreyfus, 1972, 1992). En la búsqueda de las respuestas a las cuestiones planteadas en esta sección fueron retomadas diferentes características importantes de la cibernética y la cibernética de segundo orden que habían quedado relativamente abandonadas. Ésto permitió que el estudio de la cognición tomara una nueva dirección sobre la cual se profundizará en la siguiente sección.

2.4. Enfoques alternativos

Varias propuestas se han realizado para substituir la metáfora de mente-computadora con un paradigma nuevo y diferente. Sin embargo, ha habido una tendencia para percibir este tipo de propuestas alternativas como una postura unificada de ‘oposición’ en vez de ser vistas como un conjunto diverso de paradigmas. Por ejemplo, en *Connectionism, Artificial Life, and Dynamical Systems: New approaches to old questions* (Elman, 1998) presenta tres paradigmas alternativos al computacionalismo:

Los tres enfoques comparten mucho en común. Todos reflejan un creciente interés sobre la forma en la cual, el hecho de prestar atención a los sistemas naturales (sistema nervioso, evolución, física) podría elucidar la cognición. Probablemente, ninguno de los enfoques por sí mismo está completo, pero juntos se complementan de una u otra forma que sólo podemos presagiar emocionantes descubrimientos por venir.

(Elman, 1998)

Los hechos más relevantes que menciona Elman son dos: (1) los enfoques alternativos tienden a estar dirigidos hacia sistemas biológicos y (2) existe un traslape metodológico entre los paradigmas alternativos. Sin embargo, usar los mismos métodos y tener el mismo origen no significa que exista compatibilidad entre ellos. Por lo que es importante

describir con un poco más de detalle las particularidades de cada uno de los enfoques y aclarar ciertos puntos que los diferencian unos de otros.

2.4.1. Conexionismo

El conexionismo es visto frecuentemente como la propuesta alternativa más importante a la inteligencia artificial simbólica. Básicamente propone reemplazar la metáfora mente-computadora con “un gran número de elementos de procesamiento simple llamados unidades, cada uno de los cuales manda señales excitatorias e inhibitorias hacia otras unidades” (McClelland et al., 1986). El conexionismo retoma ideas de más de un siglo de antigüedad que fueron poco más que especulaciones hasta mediados del siglo XX. A través de su trabajo sobre la estructura del sistema nervioso por el que ganó el Premio Nobel en 1906, el español Santiago Ramón y Cajal estableció la base para los estudios de redes neuronales, pero no fue hasta los años 80 que el conexionismo se convirtió en una perspectiva popular entre los científicos. Este descubrimiento no sólo se quedó en el área de las neurociencias sino que también migró a otras como la inteligencia artificial en donde generaría las Redes Neuronales Artificiales. Estas redes neuronales pueden estar en una forma no dinámica (por ejemplo, las redes *feed-forward*), en la cual sólo se hace un mapeo de las entradas y salidas como en los modelos computacionales. Aunque también existe la forma dinámica de las redes neuronales (por ejemplo, las redes recurrentes) las cuales representan sistemas dinámicos.

La teoría de Redes Neuronales Artificiales ha producido herramientas formales de gran utilidad, algoritmos de aprendizaje y representaciones para sistemas dinámicos y funciones matemáticas. Sin embargo, y pesar de las aportaciones del nuevo enfoque, el conexionismo comparte la mayor parte de los problemas asociados al paradigma computacionalista. De hecho, este es una variante del computacionalismo, no en el sentido de ‘cognición como procesamiento de información digital’ sino más bien ‘cognición como procesamiento distribuido en paralelo’. Además para poder entender la mente, la cognición y el comportamiento es necesario investigar, no solamente las entradas y las salidas, sino más bien lo que sucede en ciclos cerrados de interacciones con el mundo y de qué manera las interacciones físicas individuo-entorno se relacionan.

2.4.2. Dinamicismo

Otro enfoque alternativo y que se solidificó en los 90s fue el dinamicismo. El principal exponente de esta perspectiva van Gelder, argumenta que lo que es crucial para el área de la computación es la implementación de un procedimiento efectivo, para lo cual la característica de pasos discretos en un algoritmo juega un papel fundamental en eso. Van Gelder afirma entonces que la discretización, tanto en sus aspectos temporales como no temporales, evita que el cómputo explique muchos aspectos de la cognición, la cual afirma que es un fenómeno fundamentalmente dinámico (van Gelder, 1998). La teoría de sistemas dinámicos permite describir la evolución de los diferentes estados de un sistema a través del tiempo, incluyendo información sobre los individuos, su cuerpo,

su cerebro y su entorno, además de las interacciones que se desarrollan entre ellos. En la sección 4.2 *Teoría de sistemas dinámicos* se trata más a detalle sus características y la relación que tienen con los sistemas cognitivos.

2.5. Retos para la ciencia cognitiva

Los enfoques que se han descrito hasta este punto son alternativas que han permitido explicar algunos aspectos de fenómenos como la habilidad de resolución de problemas, el aprendizaje, el uso del lenguaje. Sin embargo, los críticos de la ciencia cognitiva han ofrecido desafíos tales como:

- Desafío respecto a las emociones: El estudio de la mente le ha dado un papel muy poco significativo a las emociones y la influencia que tienen éstas sobre la cognición.
- Desafío de la conciencia: Existen muy pocos estudios sistemáticos sobre la conciencia y los diferentes estados que puede llegar a alcanzar, así como las implicaciones que éste tiene en la mente y el pensamiento.
- Desafío del entorno: La ciencia cognitiva le ha dado un papel secundario al entorno físico y cómo influye éste en la cognición,
- Desafío del cuerpo: Similar al desafío anterior, la contribución de la corporeidad en la cognición no ha figurado en las ciencias cognitivas.
- Desafío de los sistemas dinámicos: El uso de la teoría de los sistemas dinámicos para estudiar la cognición no ha tenido una relevancia significativa.
- Desafío social: La polarización entre las contribuciones individuales y colectivas en la cognición ha representado un obstáculo para su estudio. Se siguen ignorando las aportaciones surgen a través de la interacción con más agentes y que aportan a la constitución del mismo individuo.
- Desafío matemático: Desde las matemáticas se hace una fuerte crítica al enfoque computacionalista, sugiriendo que la manera en cómo funciona el cerebro probablemente sea como una máquina cuántica.

En relación con estos puntos ha habido un gran enfoque en la generación de herramientas que permitan darle solución a los desafíos que enfrenta la ciencia cognitiva. En el siguiente capítulo se muestra una de las propuestas que ha incorporado nuevos conceptos, además del planteamiento de un marco de referencia como base para el estudio de la mente y la cognición.

Hacia una nueva ciencia cognitiva

A pesar de que la cibernética, la cibernética de segundo orden y la teoría computacional de la mente tuvieron importantes aportaciones a la ciencia cognitiva aún existían problemas que no tenían una solución práctica. Por una parte la inteligencia artificial simbólica comenzó a presentar dificultades para abordar de una manera robusta y flexible problemas del mundo real sensibles al contexto. Estas dificultades motivaron a inicios de 1990 el surgimiento de un enfoque robótico que tomaría en cuenta la cuestión corpórea y situada de los problemas (Brooks, 1991). Este nuevo enfoque tendría una gran influencia en el surgimiento de dos metodologías como el enfoque dinámico (e.g. Beer, 1995) y la robótica evolutiva (Harvey et al., 2005; Nolfi & Floreano, 2000). En la siguiente sección se mencionan características relevantes de estos nuevos enfoques y su importancia para el presente trabajo.

3.1. La inteligencia artificial corporizada-embebida

Con el fin de crear agentes artificiales que tuvieran un comportamiento mucho más flexible y robusto, la inteligencia artificial comenzó a contemplar una característica de estos agentes: el cuerpo. Para poder lograr lo anterior Brooks (1997) estableció que probablemente la razón por la que no se podía generar un comportamiento de esa índole era porque no se estaban tomando en cuenta principios de organización que pueden ser observados en organismos biológicos. Este tipo de inteligencia artificial ayudó a que se formara la cognición corporizada cuya tesis establece que muchas características de la cognición son corporizadas debido a que son profundamente dependientes de las características del cuerpo físico de un agente, de tal manera que su cuerpo -excluyendo al cerebro- juega un papel constitutivo causal de mucho peso en los procesos cognitivos del agente. Este enfoque ha generado líneas de investigación para determinar la importancia de la constitución corpórea en la cognición (cf. Pfeifer & Bongard, 2007; Shapiro, 2011; Clark, 2008).

3.2. Ciencia cognitiva enactivista

A inicios de 1990 el paradigma del enactivismo surgió originalmente como parte de la ciencia cognitiva corporizada-embebida en la publicación del libro *The Embodied Mind* (Varela et al., 1991). El enactivismo, como lo hizo la ciencia cognitiva corporizada-embebida, le otorga un papel fundamental al cuerpo y a su interacción con el ambiente. Además, concibe al agente cognitivo con un grado apropiado de coordinación con los objetos, características amenazas y oportunidades que existen en el entorno (Ward & Stapleton, 2012). El giro en la investigación de la robótica se vio sostenido por el nuevo marco de referencia propuesto por el enactivismo, sobre el cual se detallan algunos conceptos importantes.

En esta sección se describen cuestiones básicas del enfoque enactivista y piezas clave de este trabajo. Se hará particular énfasis en conceptos como autonomía y sistemas autónomos, autopoiesis, identidad y agencialidad. Finalmente, todas estas piezas son combinadas para poder mostrar de forma completa la perspectiva de la ciencia cognitiva enactivista.

3.2.1. Autonomía

El concepto de *autonomía* es el más fundamental para la ciencia cognitiva mínima, y sus orígenes surgen en con término *autopoiesis*. La noción de autopoiesis (del griego: autoproducción o autocreación) como la organización mínima de lo vivo fue propuesta por los biólogos Humberto Maturana y Francisco Varela (Maturana & Varela, 1987). Desde que el término fue acuñado en el año de 1971, sufrió ligeros cambios, pero finalmente la definición a la cual hará referencia este trabajo es la siguiente:

Un sistema autopoietico -la organización viva mínima- es aquel que produce continuamente los componentes que lo especifican, al mismo tiempo que lo hace (el sistema) como una unidad concreta en el espacio y tiempo, lo que hace posible que exista la red de componentes de producción. Haciendo una definición más precisa: un sistema autopoietico es organizado (definido como unidad) como una red de procesos de producción (síntesis y destrucción) de componentes de tal manera que estos componentes: 1. se regeneran continuamente y realizan la red que los produce, y 2. constituyen el sistema como una unidad distinguible en el dominio en el cual existen.

(Varela, 1997, p. 75)

Sin embargo, el uso del concepto de autopoiesis es asociado comúnmente con la autoproducción de sustancias químicas en el organismo. No obstante, la autopoiesis no es la única forma que un sistema autónomo puede tomar, por lo que Varela propone características operacionales más generales para la autonomía en sistemas vivos mínimos:

Vamos a decir que los sistemas autónomos son cerrados organizacionalmente. ésto es, su organización está caracterizada por procesos de tal manera que:

1. los procesos están relacionados como una red, por lo que dependen recursivamente uno del otro en la generación y realización de los procesos mismos, y
2. ellos constituyen al sistema como una unidad reconocible en el espacio (dominio) en el que los procesos existen.

(Varela, 1979, p. 55)

La noción de autonomía como cerradura organizacional permite que se pueda aplicar no solamente a sistemas vivos, sino también a otro tipo de sistemas como el sistema inmune, el sistema nervioso, e incluso sistemas sociales (Varela, 1991). Sin la autonomía proporcionada por la cerradura organizacional, el sistema sería incapaz de definir su propia identidad como individuo por lo que seguiría siendo una colección de componentes definida externamente y que se ha elegido designar como ‘agente’ sólo por convención.

3.2.2. Identidad

El problema de la constitución de identidad ha tomado un importante lugar en la robótica tradicional y en la inteligencia artificial. Ésto se debe a las decisiones arbitrarias que los investigadores se ven forzados a hacer cuando se trata de distinguir el sistema de su ambiente. No obstante, la característica de autonomía en un sistema proporciona una identidad esencialmente autoconstituida, ya que su actividad generativa marca lo que puede ser contado como parte del sistema y lo que pertenece al entorno.

Como resultado, la identidad se denota como una red cohesiva de estructuras dinámicas que puede ser distinguida del entorno; además de que aparece un grupo de correlaciones altamente integradas, donde esas correlaciones son el efecto recursivo de la actividad de un conjunto de variables. A ese grupo de correlaciones se le llama identidad porque define al sistema como una unidad autosustentable (Barandiaran & Moreno, 2006).

3.2.3. Agencialidad

El término agencialidad se refiere a la habilidad que tiene un sistema autónomo de alcanzar un estado de adaptación no solamente por medio de la reorganización interna, sino también por la regulación adaptada de sus interacciones sensomotoras. La noción de agencialidad es introducida como la forma más básica de existencia autónoma que puede convertirse en parte de un sistema multi-agente, es decir, un sistema en el cual la dinámica relacional de las interacciones interindividuales pueden asumir una organización autónoma. A ésto le sigue una consideración de las condiciones necesarias

para la interacción social, que requiere una forma más específica de agencia (mentalidad) capaz de constituir un dominio cognitivo.

3.3. Dimensiones de la corporización

Esta línea de desarrollo que hubo en la ciencia cognitiva permitió aportar al estudio de la cognición cómo es que la actividad corporal y las características del cuerpo como tal se relacionan con la cognición del individuo. Existen tres características importantes por medio de las cuales se relacionan el cuerpo con la cognición: la autorregulación corporal, el acoplamiento sensoriomotor y la interacción intersubjetiva (Thompson & Varela, 2001).

3.3.1. Autorregulación corporal

Esta característica se refiere a que un agente artificial tenga la capacidad de regular sus componentes internos de tal forma que sean modificados con el fin de mantener una interacción estable con el ambiente. Por ejemplo, la capacidad de autorregulación le da un comportamiento robusto y flexible al agente, ya que éste tiene la capacidad de poder adaptarse a los cambios constantes del entorno. La base de esta autorregulación es el sistema nervioso autónomo que poseen los organismos, de tal manera que los sensores y efectores de éste están relacionados con procesos neuronales en el cerebro.

Los sistemas de inteligencia artificial predominantes no cuentan con esta capacidad de adaptabilidad ya que a lo más que pueden aspirar es a las propiedades que el creador/diseñador le otorgue. Una manera de generar agentes artificiales que no sean tan dependientes de la persona que los construye es por medio de la metodología de robótica evolutiva, ya que permite que la creación de “individuos” sea definida por medio de un algoritmo evolutivo.

3.3.2. Acoplamiento sensoriomotor

Como ya se había mencionado anteriormente, el enfoque enactivista toma en consideración las interacciones que existen entre el individuo y su entorno. El comportamiento de un agente en determinado escenario puede ser visto como un ciclo de acoplamiento sensoriomotor con el ambiente, es decir, lo que el agente sensa está en función de sus movimientos y a su vez esos movimientos están determinados por la información recibida por los sensores (O'Regan & Noë, 2001). La idea de este nuevo enfoque corporizado-situado es que la cognición necesita de la exploración en el mundo la cual es mediada a través de contingencias sensoriomotoras.

3.3.3. Interacción intersubjetiva

Otra de las cuestiones importantes del enactivismo es que no sólo se toman en consideración las interacciones del agente con el entorno, sino también las interacciones con otros agentes. Di Paolo, Rohde y De Jaegher (2007) establecen de una forma mucho más clara la idea anterior, sugiriendo que la investigación enfocada en las características corpóreas de los agentes necesita ser complementada con el estudio de los procesos interactivos entre más de un agente. Lo anterior permite extender la investigación de lo individual a lo colectivo, incursionando en el área de la cognición social. El estudio del comportamiento de un conjunto de agentes y sus interacciones podría ayudar a resolver problemas que de manera individual han sido difíciles de abordar.

En términos generales, la manera que tiene el enactivismo de abordar el análisis de los sistemas es a través de un enfoque dinámico en el que el sistema nervioso del agente, su cuerpo y el ambiente que lo rodea son vistos como un sistema dinámico acoplado (Fig. 3.1). De tal manera que el resultado completo del sistema es producto tanto de las aportaciones de cada elemento como en general de las interacciones que existen entre ellos.

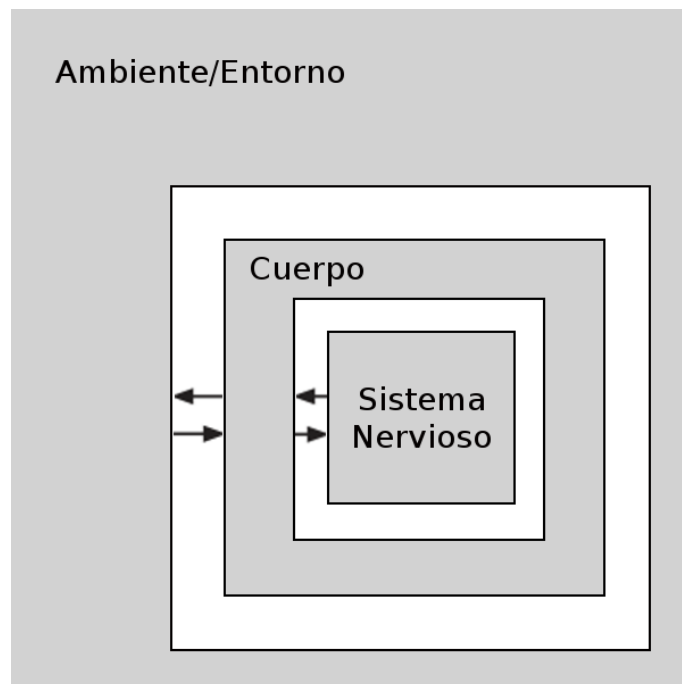


Figura 3.1: Perspectiva dinámica de agentes corporizados situados. El sistema nervioso, el cuerpo y el ambiente son conceptualizados como sistemas dinámicos que se encuentran en constante interacción entre ellos. Basada en la figura original de Beer (2000).

El diseñar sistemas artificiales que cumplan con las características antes enlistadas

3. HACIA UNA NUEVA CIENCIA COGNITIVA

ha implicado retos de diferente índole. Específicamente, para la inteligencia artificial el reto sería implementar sistemas que incluyan procesos de interacción corporizados y situados, algo opuesto a las herramientas tradicionales de ingeniería e inteligencia artificial.

El planteamiento de la presente tesis está basado en el contexto de estas recientes líneas de investigación. El objetivo principal es estudiar y analizar a través de la metodología de la robótica evolutiva los mecanismos de interacción y cognición social entre agentes artificiales. La siguiente sección describe más a detalle el uso de la robótica evolutiva para el diseño de los agentes y los sistemas dinámicos para el análisis del comportamiento y los procesos de interacción.

Metodologías y herramientas de estudio

Con el fin de poder construir sistemas mucho más flexibles y robustos, agentes artificiales autónomos situados se han venido desarrollando empleando diferentes metodologías enfocadas en esos aspectos. Los robots industriales son un ejemplo de robots que aún no se convierten en robots adaptables: ellos ejecutan la misma secuencia de acciones repetidamente. La idea básicamente es poder crear agentes que tengan la capacidad de actuar en entornos complejos y con ruido donde los mecanismos subyacentes de dichas capacidades generen un comportamiento adaptativo en el agente. En esta búsqueda de nuevas estrategias para desarrollar este tipo de proyectos la tendencia ha sido voltear la mirada hacia los sistemas y organismos biológicos y así obtener información que pueda ser llevada a la práctica exitosamente. Una de las metodologías que ha concentrado sus esfuerzos en generar este tipo de sistemas es, como ya se había mencionado anteriormente, la robótica evolutiva, sobre la cual está basado el presente trabajo. En las siguientes secciones se hablará sobre los orígenes de la robótica evolutiva, su utilización en modelos mínimos de cognición social y su importancia en este trabajo de investigación.

Como parte de la metodología de robótica evolutiva se implementó una red neuronal recurrente de tiempo continuo de la cual se explica su modelo matemático, estructura y características principales que fueron determinantes para construir el modelo de cognición social utilizado en esta tesis. Además, se presenta una introducción de la teoría de sistemas dinámicos utilizada para comprender mejor el comportamiento de los agentes.

4.1. Robótica evolutiva

Como parte del diseño de robots existen elementos que deben ser considerados simultáneamente: la arquitectura de control, el sistema motor, los sensores, su morfología, etc. (Siciliano & Khatib, 2008). Uno de los grandes retos de la robótica es que todos estos elementos interactúen y de manera conjunta determinen el comportamiento del robot. El concepto de inteligencia artificial corporizada (Pfeifer & Bongard, 2007) es un punto de vista alternativo en el cual el robot, su cuerpo, el ambiente y las interacciones

de sus componentes son estudiados como un todo.

4.1.1. ¿Qué es?

La robótica evolutiva es una metodología que emplea algoritmos evolutivos para generar robots adaptables al entorno a través de procesos inspirados en la evolución natural. Los algoritmos evolutivos han sido utilizados para diseñar distintas partes de un robot: el sistema de control, su estructura morfológica, la parte sensomotora, sólo por mencionar algunas. La robótica evolutiva enfatiza la característica corporizada y situada de los sistemas, además de la interacción cerrada que mantiene el cerebro, el cuerpo y el ambiente, lo que es crucial para el surgimiento de inteligencia, comportamiento adaptativo y procesos cognitivos (e.g Clark, 1997; Chiel & Beer, 1997; Nolfi & Floreano, 2002).

La robótica evolutiva se caracteriza por ser un enfoque muy general, ya que permite la síntesis de comportamiento tanto de control como de morfología, especificando solamente la tarea que se necesita realizar. Las técnicas de robótica evolutiva tienen dos modos de aplicación: *offline* (no son en tiempo real) y *online* (en tiempo real). En el caso del primer tipo de aplicaciones se generan las simulaciones en donde los controladores son evolucionados una determinada cantidad de generaciones o hasta que se alcance un criterio en específico. El segundo tipo de aplicaciones se presentan cuando los robots están realizando una tarea y al mismo tiempo son ejecutados algoritmos genéticos para ejecutar correctamente la tarea o incluso mejorar su desempeño (Silva et al., 2016). En general, las aplicaciones que son más utilizadas en la práctica son las que no son en tiempo real. Por otro lado, los investigadores que aplican las técnicas de robótica evolutiva generalmente explotan dos características muy importantes: la auto-organización del sistema y la optimización holística, lo que permite eliminar la necesidad de hacer especificaciones muy detalladas manualmente, ya que esa tarea es resultado del algoritmo evolutivo.

Por otro lado, debido a los requerimientos que tienen que ser cubiertos en la robótica evolutiva, se ha descubierto que los controladores que más se han ajustado son los basados en redes neuronales. Este tipo de controladores son utilizados para encontrar ya sea los mejores parámetros o la mejor arquitectura del controlador. Existen muchos tipos de redes neuronales artificiales, además de distintas formas de evolucionarlas, sin embargo, el tipo de red neuronal más utilizado en la robótica evolutiva es la red neuronal recurrente de tiempo continuo (CTRNN por sus siglas en inglés). Así mismo, este tipo de redes neuronales tienen la particularidad de ser las más adecuadas para representar sistemas dinámicos. En la siguiente sección se habla más a detalle sobre este tipo de redes neuronales y sus características.

4.2. Redes neuronales artificiales

Las redes neuronales artificiales consisten en una gran colección de unidades neuronales (también conocidas como neuronas artificiales) que modelan la manera en la que un cerebro biológico resuelve problemas. Lo anterior se realiza a través de grandes grupos de neuronas biológicas conectadas por medio de axones. Cada unidad neuronal está conectada con muchas otras por medio de enlaces que pueden tener un efecto excitatorio o inhibitorio en la conexión con las otras neuronas.

La manera en que una red neuronal artificial está interconectada se llama topología, arquitectura o grafo. El hecho de que la interconexión pueda ser realizada de diferentes maneras hace que existan distintas topologías, las cuales son divididas básicamente en dos clases (Fig. 4.1).

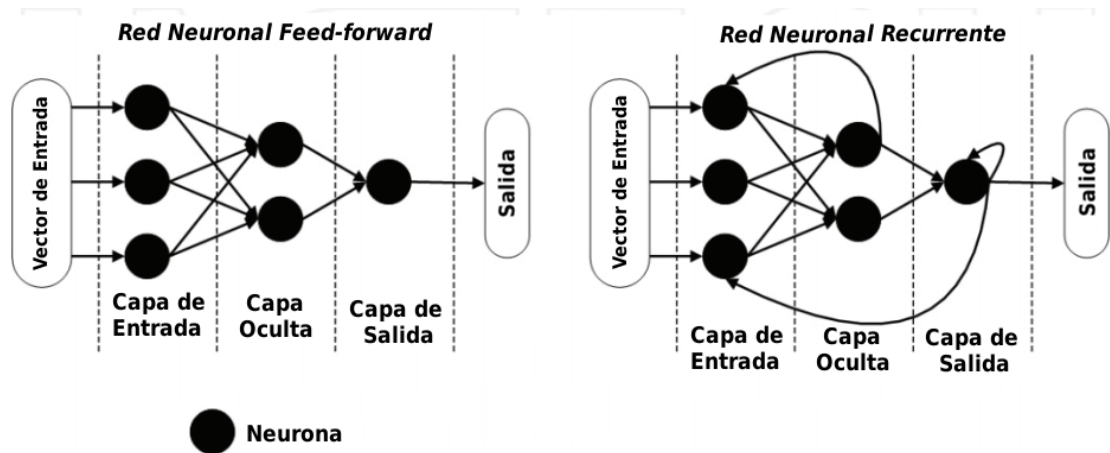


Figura 4.1: Principales topologías de las redes neuronales artificiales.

4.2.1. Redes Neuronales Artificiales *feed-forward*

La característica principal de las redes *feed-forward* es que la información procesada debe de fluir en una sola dirección partiendo de las entradas hacia la salida sin ciclos que vayan hacia atrás. No existe límite para el número de capas, para el tipo de función de transferencia utilizado en una neurona artificial o el número de conexiones entre las neuronas artificiales. La red neuronal artificial *feed-forward* más sencilla consiste en un solo perceptrón que solamente es capaz de aprender problemas de separación lineal. La figura 4.2 muestra un ejemplo de una red neuronal artificial *feed-forward*.

En resumen, las características principales de este tipo de redes son:

- La activación de las neuronas se transmite de la entrada a la salida a través de “capas ocultas”.

- Matemáticamente implementan asignaciones estáticas de entrada-salida.
- El algoritmo de entrenamiento más popular es el supervisado.
- Han demostrado ser muy útiles en muchas aplicaciones prácticas como aproximaciones de funciones no lineales y como clasificación de patrones.

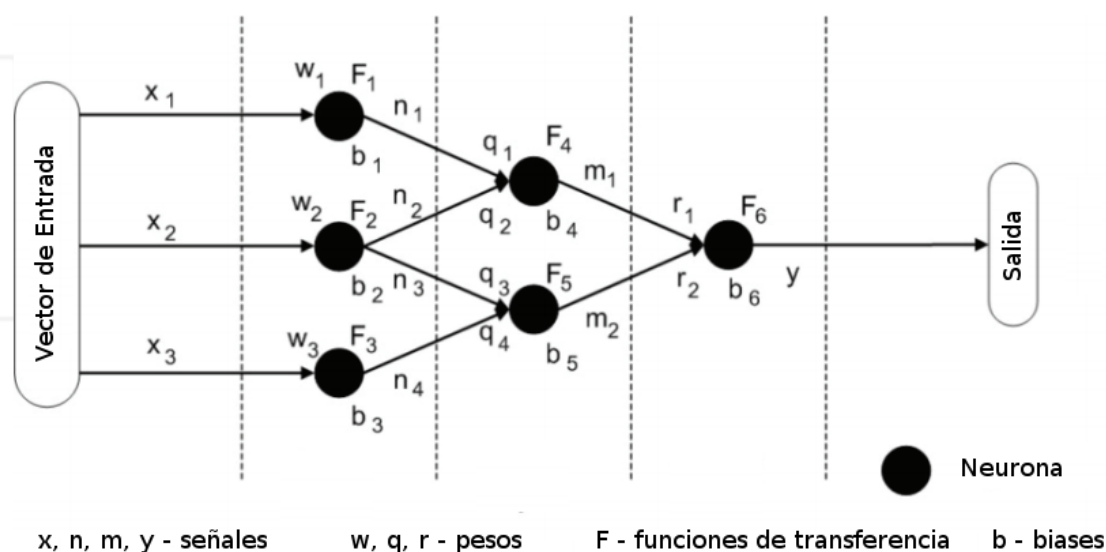


Figura 4.2

4.2.2. Redes Neuronales Recurrentes

Este tipo de redes, también conocidas como RNN por sus siglas en inglés, son parecidas a las *feed-forward* solamente que no existe la limitación respecto a los ciclos hacia atrás. Es decir, la información no sólo se transmite de la entrada a la salida sino también en sentido inverso. Esta característica introduce la noción de tiempo al modelo y debido a esto, se crea un estado interno en la red que permite exhibir un comportamiento temporal dinámico. A diferencia de las redes *feed-forward*, las RNN pueden usar su memoria interna para procesar secuencias arbitrarias de entrada. El tipo de recurrencia autoreferenciada que existe en una neurona se hace a través del tiempo, es decir, al tiempo t los nodos con enlaces recurrentes reciben como entrada los datos actuales del punto $x^{(t)}$ y también los valores de los nodos ocultos $h^{(t-1)}$ del estado previo de la red. La salida $\hat{y}^{(t)}$ en cada tiempo t es calculada dados los valores $h^{(t)}$ de los nodos ocultos en el tiempo t . La entrada $x^{(t-1)}$ en el tiempo $t - 1$ puede influenciar la salida $\hat{y}^{(t)}$ en el tiempo t y también en los tiempos sucesivos a través de las conexiones recurrentes.

Existen otras RNN como las redes de Hopfield, Elman, Jordan, bidireccionales y otras más son casos especiales de redes recurrentes. Enlistando las características básicas de las RNN tenemos lo siguiente:

- Las RNN tienen por lo menos un ciclo de conexión sináptica.
- Todas las redes neuronales biológicas son recurrentes.
- Matemáticamente implementan sistemas dinámicos.
- Su dinámica no lineal les permite actualizar sus estados ocultos de maneras complejas.
- Se conocen varios tipos de entrenamiento, no existe la clara predominancia de alguno en especial.

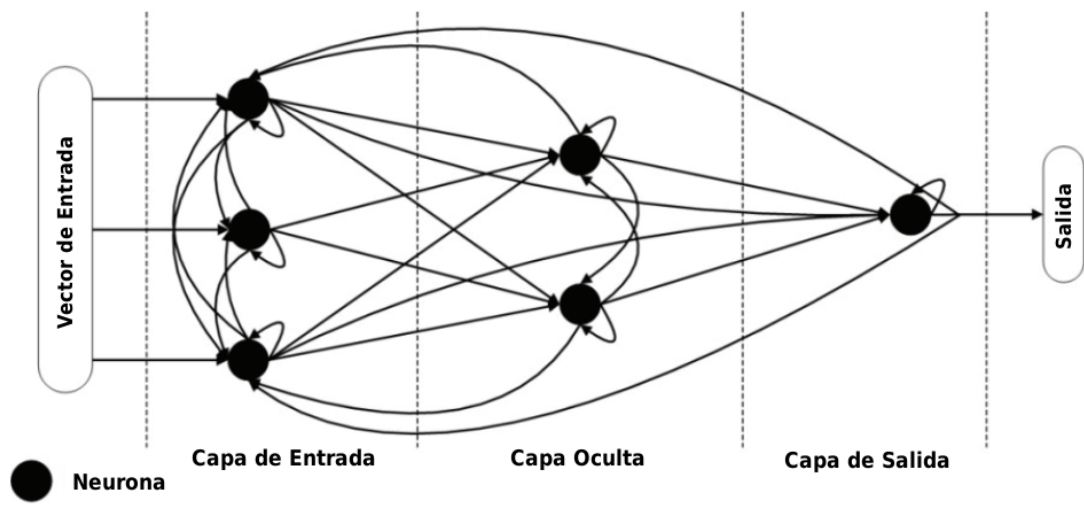


Figura 4.3

En la siguiente sección se hace una particular descripción de un tipo de RNN: las redes neuronales recurrentes de tiempo continuo.

4.3. Redes Neuronales Recurrentes de Tiempo Continuo

La inteligencia artificial simbólica enfatiza los comportamientos de bajo nivel en el corazón de las estrategias de supervivencia de todas las especies. Los primeros trabajos en cibernética realizados por Gray-Walter establecieron un contexto experimental en el que los robots con ruedas, que contenían sensores y motores conectados por circuitos analógicos simples, podían diseñarse para producir un comportamiento visiblemente realista. Desde el desarrollo de algoritmos genéticos y el incremento de capacidad

computacional ha sido posible desarrollar algoritmos compactos que han permitido que un robot físico pueda desarrollar determinadas tareas cognitivas.

En tiempos recientes se ha hecho un gran esfuerzo para comprender cómo los algoritmos de inspiración biológica extremadamente sencillos pueden aprender tareas como reconocimiento de objetos, atención selectiva, memoria simple e interacción entre individuos. Una de las herramientas que se han utilizado para tales propósitos son las redes neuronales recurrentes de tiempo continuo (CTRNN por sus siglas en inglés), las cuales se encuentran *corporizadas* en los agentes simulados y *situadas* en entornos físicos simples.

4.3.1. Descripción técnica de las CTRNN

Las CTRNN son un tipo de red neuronal artificial para modelar neuronas que se comportan de manera general bajo el siguiente modelo matemático:

$$\tau_i \dot{y}_i = -y_i + \sum_{j=1}^N w_{ji} \sigma(g_i(y_j + \theta_j)) + I_i \quad i = 1, \dots, N \quad (4.1)$$

donde y es el estado de cada neurona, τ es su constante de tiempo, w_{ji} es la fuerza de la conexión que va de la neurona j a la neurona i , g_i es un factor de ganancia para cada una de las neuronas, θ representa el sesgo, $\sigma(x) = 1/(1 + e^{-x})$ es la función de activación logística estándar, y por último la I representa las entradas de la red neuronal (Beer, 1995).

A diferencia de otro tipo de neuronas, como por ejemplo las prealimentadas o también conocidas como *feedforward*, las CTRNNs permiten generar un comportamiento que está basado en estados previos del sistema y que incluso le permite anticiparse a eventos futuros. Además el agente artificial puede desarrollar la capacidad de iniciar una acción independientemente de la situación inmediata que haya ocurrido. En contraste con las redes *feedforward* que sólo permiten generar comportamientos reactivos.

Las CTRNN permiten la recurrencia y como ya se había mencionado antes, esto significa que las conexiones de la red pueden existir en cualquier dirección, incluyendo posibles conexiones de cualquier nodo a sí mismo. La combinación de la recurrencia y el estado interno hace que el sistema tenga la capacidad producir patrones internos complejos y que a su vez éstos posean un sistema de respuesta basado en la memoria de hechos anteriores ocurridos en el ambiente. Cada una de las neuronas tiene tres parámetros -constante de tiempo, sesgo, ganancias de entrada y salida- asociados con su comportamiento, mientras que cada conexión entre los nodos tiene un parámetro -el peso sináptico. Debido a la compleja relación entre los parámetros de la red y el comportamiento de la misma, aunado a que no se hace una especificación explícita de los métodos de solución para desarrollar determinada tarea, el método más común para que una CTRNN tenga cierto comportamiento es el uso de algoritmos genéticos.

Por otro lado, las CTRNN pueden ser vistas como un sistema dinámico típico debido a que: tienen un número arbitrario de variables que cambian en relación al tiempo

bajo ciertas reglas dependiendo de los valores actuales de esas mismas variables. Las CTRNN no pueden ser consideradas como cualquier otra red neuronal artificial, ya que si pensamos en esta última, lo más inmediato es relacionar el aprendizaje de la misma con el cambio en el peso de la conexión entre sus neuronas. A diferencia de las CTRNN en las que el hecho de que sus pesos sinápticos no se muevan no implica que no puedan aprender (Izquierdo et al., 2008).

4.3.2. Comportamiento de las CTRNN

El especial interés por las CTRNN como controladores de agentes artificiales con comportamiento mínimo cognitivo proviene de la gama potencialmente ilimitada de la dinámica temporal de la que son capaces. La interacción de la CTRNN con el entorno consiste en vectores de entradas y salidas de valor real, actualizándose continuamente y con frecuencia. Beer (Beer, 1995) proporciona un análisis matemático extenso del comportamiento de CTRNN pequeñas. Una noción clave en dicho análisis es que las neuronas con autoconexiones adecuadamente fuertes se pueden saturar, lo que significa que su propia retroalimentación domina su entrada y los bloquea en un determinado estado. De esta manera, los nodos pueden actuar como conmutadores internos que influyen en el comportamiento del resto de la red. Tales nodos pueden invertir estados. Más ampliamente, debido a su estado interno y a su recurrencia, el estado de una CTRNN en el instante t es determinado no sólo por la entrada actual, sino también por la historia de las entradas hasta el tiempo t , y el estado inicial de la red.

Las características tanto técnicas como de comportamiento de las CTRNN fueron importantes en el diseño e implementación del modelo de robótica evolutiva del presente trabajo. Como se verá más adelante, se parte de la posición de considerar a los sistemas cognitivos como sistemas dinámicos y por tanto la implementación de CTRNN se vuelve fundamental para la generación de modelos cognitivos mínimos. Además de la noción de tiempo que se considera imprescindible en este tipo de sistemas.

4.4. Algoritmos genéticos

Para lograr generar controladores que tengan la capacidad de adaptabilidad con su entorno ha sido necesaria la implementación de algoritmos genéticos. Un algoritmo genético es una técnica de búsqueda inspirada en los procesos de evolución natural (Goldberg, 1989; Holland, 1975). Los algoritmos genéticos han demostrado ser poderosas técnicas donde el espacio de búsqueda es grande, en búsquedas complejas, multimodales, con ruido, etc. Un algoritmo genético básico opera de la siguiente manera. El espacio de búsqueda puede ser codificado de diferentes formas: binaria, entera o real. Se comienza con una población de individuos aleatoria donde sus genes contienen los parámetros que necesitan ser optimizados. En cada una de las iteraciones el desempeño de cada individuo con respecto a la tarea objetivo es evaluado. Una siguiente generación de individuos es producida aplicando un conjunto de operadores genéticos sobre los in-

dividuos seleccionados de la generación previa. Los individuos son seleccionados para reproducirse con una probabilidad proporcional a la aptitud que tuvieron para resolver la tarea objetivo. Los operadores genéticos estándar son la *mutación* (en donde los bits son aleatoriamente alterados) *cruza* (en la cual ciertas partes de los genes de los individuos son intercambiadas). La *cruza* es generalmente considerada como el principal mecanismo de búsqueda, dejando a la *mutación* como un operador cuyo rol es el de mantener la diversidad en la población y además asegurar que cada punto del espacio de búsqueda tiene alguna probabilidad de ser visitado. Por medio de la iteración de los procesos de selección, *cruza* y *mutación*, la población acumula información acerca de la distribución de los valores de aptitud en el espacio de búsqueda. Esta información permite que las búsquedas siguientes se enfoquen en subespacios que ayuden a alcanzar el valor de aptitud esperado.

Ha inicios de los 90 hubo un gran interés por combinar las redes neuronales y los algoritmos genéticos. Principalmente, existen cuatro enfoques básicos: (1) utilizar un algoritmo genético para encontrar los pesos de conexión adecuados en vez de emplear retropropagación; (2) usar algoritmos genéticos para encontrar buenos parámetros de configuración; (3) usar un algoritmo genético para encontrar buenos pesos iniciales para la retropropagación; y (4) usar un algoritmo genético para encontrar buenas arquitecturas *feedforward* para retropropagación. El trabajo pionero en aplicar los algoritmos genéticos para evolucionar CTRNN fue hecho por Randall Beer y John Gallagher en su artículo *Evolving Dynamical Neural Networks for Adaptive Behavior* (Beer & Gallagher, 1992). Justo esta combinación fue implementada para poder crear los controladores de los agentes artificiales del modelo presentado en este trabajo. En el capítulo 5 se presenta una descripción detallada de las características tanto del algoritmo genético como de la CTRNN implementados.

4.5. Teoría de sistemas dinámicos

4.5.1. Conceptos

Con el fin de tener dar una explicación introductoria a la teoría de sistemas dinámicos es necesario comenzar con la descripción de algunos conceptos claves. En primer lugar se entiende como un *sistema* a un conjunto de componentes y sus relaciones, y que percibimos como una sola unidad. Específicamente, los *sistemas dinámicos* (que son un tipo de sistemas) presentan cambios que ocurren a través del tiempo; por tanto, el estado de un sistema es la descripción del mismo en un tiempo determinado.

Una parte importante para poder entender un sistema es a través de las variables que lo componen. Eso no implica que se pueda hacer una descripción totalmente completa y detallada, sin embargo, las variables deberán describir completamente el estado del sistema en términos matemáticos. Las variables que describen el estado del sistema dinámico se llaman *variables de estado*. El conjunto de todos los posibles valores de las variables de estado es llamado *espacio de estado*. Existen dos tipos de espacio de

estado: los discretos y los continuos. Los discretos solamente toman valores enteros, mientras que los continuos toman valores reales. Para este último caso existe un nombre en particular: *estado de fases*. Una característica especial de los sistemas dinámicos es que son sistemas determinados por los estados. Es decir, un atributo de un sistema puede ser predicho o entendido por medio del estado que ocupa actualmente. Los sistemas determinados por los estados se pueden describir en términos de las transiciones entre los estados en un instante de tiempo y el siguiente. Específicamente, la herramienta matemática que se utiliza principalmente para describir la tasa de cambio son las ecuaciones diferenciales.

Más formalmente, los sistemas dinámicos son una rama de las matemáticas que estudia el comportamiento de sistemas complejos¹ por medio del uso de ecuaciones diferenciales. Visto de otra forma, trata de entender o por lo menos describir los diferentes estados por los que pasa un sistema, incluyendo los cambios dentro de sus componentes y cómo son afectados unos a otros a través del tiempo bajo determinadas reglas (Beer, 2000).

Pero, ¿cómo se relacionan los sistemas dinámicos con la creación de agentes artificiales autónomos y el estudio de la cognición? Como se mencionó en el capítulo anterior una posibilidad para crear agentes lo suficientemente robustos, flexibles y autónomos es retomar características de organismos biológicos que existen en la naturaleza. La importancia que tiene lo anterior, es que los fenómenos que ocurren en la naturaleza se desenvuelven a diferentes escalas en tiempo real. Muchos fenómenos no ocurrirían si no sucedieran en el momento indicado. Los sistemas dinámicos permiten ser analizados en diferentes escalas de tiempo dependiendo qué tan denso sea el espacio de fases, es decir, dados dos puntos en el espacio de fase, podemos encontrar más puntos entre esos dos, con lo cual se puede analizar el sistema en cualquier instante de tiempo (van Gelder & Port, 1995). Así mismo, uno de los objetivos en el presente trabajo es estudiar el proceso de cognición social en agentes artificiales, y como será profundizado en las siguientes secciones, al ser la cognición social un fenómeno de la naturaleza, ésta puede ser estudiada y simulada utilizando sistemas dinámicos.

4.5.2. El enfoque de los sistemas dinámicos

Los enfoques clásicos para entender y estudiar la cognición, como el computacionalista, generan demasiadas dualidades durante el proceso de análisis y no se enfocan en lo que realmente ha sido el objetivo inicial: entender cómo funciona la cognición. La

¹Área que se refiere al estudio e investigación de las interacciones de las partes de un sistema y cómo éstas dan lugar a comportamientos colectivos, además de su interacción con el entorno. (Bar-Yam, 2002) El comportamiento global de estos sistemas no es fácilmente predecible conociendo sólo las características de las partes individuales. Es común que para su representación sean utilizadas redes donde los nodos representan los componentes y los enlaces las interacciones. Ejemplos de sistemas complejos son: el cerebro, organismos biológicos, el clima, la economía, ecosistemas, etc.

perspectiva dinámica asume dos partes muy importantes. Por un lado la continuidad, que permite ver al sistema como un ensamblaje que se desenvuelve a través del tiempo y en función de su historia. Y por otro lado, está el acoplamiento, lo que indica que los componentes se encuentran entrelazados y que interactúan entre sí a diferentes niveles: entre el cerebro y el cuerpo del individuo, y entre el cuerpo y el mundo (Thelen & Smith, 1998).

4.5.3. Sistemas cognitivos como sistemas dinámicos

Los dinamicistas consideran a la teoría de sistemas dinámicos como la herramienta más apropiada para el modelado y estudio de la cognición. La teoría de sistemas permite modelar interacciones complejas entre individuos, así como con su ambiente, sin la necesidad de reducir el fenómeno a niveles de cognición individual.

Timothy van Gelder y Robert F. Port formalizaron la idea anterior postulando la llamada Hipótesis Dinámica que dice:

Los sistemas cognitivos naturales son sistemas dinámicos, y son mejor entendidos desde la perspectiva dinámica. Como su contraparte computacional, la Hipótesis Dinámica forma un marco de referencia general dentro del cual se pueden construir teorías detalladas de aspectos particulares de la cognición. Sólo sabremos que la Hipótesis Dinámica es verdadera si, con el paso del tiempo, las mejores teorías de procesos cognitivos son expresadas en términos dinámicos.

(van Gelder & Port, 1995, p. 5)

Contrastando diversos puntos entre el enfoque computacional y el enfoque dinámico, en el primero se le da mucha relevancia a las entradas y las salidas del sistema, o lo que sería el punto inicial y el final; mientras que en el enfoque dinámico se analiza cómo va evolucionando el comportamiento del sistema a través del tiempo. En esta misma línea, el enfoque computacional considera que los símbolos que forman parte del sistema no cambian de un momento a otro, y desde la perspectiva dinámica ese cambio es parte del sistema mismo y por lo tanto es contemplado.

Algo que es importante enfatizar es que el enfoque dinámico no es algo totalmente opuesto a otros enfoques, sino simplemente es una manera de unificar los distintos hilos que constituyen a la cognición. Lo que aporta la perspectiva dinámica es permitir entender los procesos cognitivos como un sistema complejo de dinámicas entrelazadas y anidadas y de interacciones auto-organizantes en muchos niveles de análisis, incluyendo aquéllas entre el cerebro y el cuerpo, y entre el cuerpo y el mundo exterior.

Modelos Mínimos de Comportamiento

Cognitivo

La teoría de sistemas dinámicos ha ofrecido importantes herramientas para el estudio y entendimiento de algunos aspectos sobre “cómo funcionan los agentes cognitivos”, sin embargo algo que aún no ha sido estudiado en profundidad es “qué es lo que hace que algo sea cognitivo”. Como ya se vio anteriormente, diferentes enfoques no han sido suficientes para poder responder o explicar este tipo de fenómenos.

Alrededor del año de 1990, comenzó un fuerte desarrollo en la investigación dentro del área de la inteligencia artificial utilizando evolución artificial. En sus inicios, esta investigación se dio principalmente en el área de la robótica. Sin embargo, el uso de la robótica no sólo se ha enfocado en la construcción de robots físicos, sino también ha habido un gran interés para el estudio y entendimiento de la cognición, ya que esta herramienta se adecúa muy bien para la generación de modelos de agentes artificiales o sintéticos. En este capítulo se habla con detalle sobre esta metodología y sus aplicaciones. En las primeras secciones se presenta un contexto más amplio de cognición, así como también modelos mínimos de la misma. En las secciones posteriores se particulariza sobre el paradigma del cruce perceptual para estudiar la cognición y posteriormente se ahonda sobre los modelos sintéticos que han tomado como inspiración este paradigma.

5.1. Cognición Social

Los organismos que son distintivamente cognitivos parecen estar dotados de un subsistema específico que está desacoplado jerárquicamente de los procesos de construcción metabólica: el sistema nervioso. Debido a que el comportamiento es controlado por la dinámica neuronal en organismos dotados de sistema nervioso, el nivel de modelación de las interacciones de comportamiento no es el de los ciclos metabólicos, sino el de un

dominio dinámico diferente constituido por actividad neural y su acoplamiento corporizado sensomotor con el ambiente.

Pero, ¿qué es lo que hace que algo sea cognitivo? La red adaptativa de dependencias dinámicas que son creadas dentro y entre el sistema nervioso y su acoplamiento con el cuerpo metabólico y con el entorno será llamado *organización cognitiva*. Ésto es diferente al comportamiento adaptativo genérico, el cual sólo necesita satisfacer restricciones externas ya definidas. Por externas se hace referencia a que las restricciones de viabilidad que el comportamiento adaptativo debe satisfacer son externas al comportamiento y la organización sensomotora que las genera.

Partiendo de las definiciones conceptuales descritas en secciones anteriores se tomará como *cognición* a la regulación compensatoria de una red de dependencias estables entre la creación de estructuras sensomotoras y su preservación durante un proceso histórico/de desarrollo. Esta definición le da a la cognición la característica de autonomía adaptativa en el dominio corporizado y situacional. Aunado a ésto, y algo que es muy importante para los procesos cognitivos son las acciones que ejecutamos: las actividades que hacemos están basadas en información recibida previamente, y a su vez esas actividades determinarán la nueva información que se reciba posteriormente. Lo anterior indica claramente un ciclo sensomotor que le da relevancia a las características del entorno, ya que eso influye en lo que se percibe y en lo que se ejecuta.

Debido a que la cognición se ve impactada por los elementos de su entorno es necesario también contemplar la interacción que existe con otros individuos. Esa interacción con otros individuos es llamada *interacción social*, la cual se da cuando dos individuos pueden establecer un acoplamiento co-regulado, cuidando que conserven su propia autonomía durante la interacción, además de generar una nueva autonomía dinámica. Por tanto, conjuntando la cognición y la parte social, la cognición social es utilizada para describir los fenómenos cognitivos que surgen o están presentes en la interacción de dos o más individuos. Además, la cognición social permite estudiar cómo es posible: generar interacciones y mantenerlas, formar relaciones con otros individuos/agentes, entender a otra individuo, actuar conjuntamente para cumplir un objetivo, etc.

Una parte importante de la cognición social son las interacciones sociales, éstas están caracterizadas por involucrar diferentes tipos de dimensiones tanto verbales como no verbales, presentar variaciones de contexto, en ocasiones contar con un medio como la tecnología, tener una actividad recíproca y conjunta, ocurrir a diferentes escalas de tiempo y ser robustas a interrupciones externas (De Jaegher et al., 2010). Un caso particular donde interviene la cognición social es el de cruce de percepciones entre dos agentes. En la siguiente sección se describe las características de este fenómeno y cómo es que ayuda al estudio de pruebas de concepto relacionadas a la cognición social.

5.2. El paradigma del cruce perceptual

El cruce perceptual entre dos agentes es la acción recíproca de percibir al otro mientras se es percibido. ¿Cómo es que en la vida real o a través del uso de dispositivos

tecnológicos es posible reconocer la presencia de otras personas y bajo qué condiciones éstas pueden ser distinguidas del resto de los objetos del mundo? Con el fin de abordar esta pregunta se comenzaron a explorar las condiciones básicas necesarias para que ciertos individuos pudieran reconocer la presencia de otra persona en una situación de interacción. Algo que es importante descubrir es: ¿qué propiedades se pueden atribuir a los individuos y cuáles son parte del fenómeno colectivo por sí mismo?

Con el fin de analizar las interacciones sociales entre las personas han surgido diferentes enfoques, algunos de ellos interesados en las interacciones de percepción en un entorno social. Por ejemplo, el enfoque interaccionista argumenta que el reconocimiento de intencionalidad en otra persona es intrínseco a la actividad perceptual compartida (cf. Di Paolo & De Jaegher, 2012; De Jaegher et al., 2016). Cuando se presenta el escenario donde existe una percepción recíproca entre dos objetos intencionales las interacciones sociales emergen de las dinámicas entre los sujetos como resultado del proceso de interacción por sí mismo (e.g. Butterworth & Jarrett, 1991). Relacionado a este tema, uno de los experimentos más destacados como enfoque alternativo para la psicología social es el experimento minimalista de cruce perceptual desarrollado por Auvray et al. (2009).

El experimento del cruce perceptual es un intento por investigar las condiciones más básicas involucradas en el reconocimiento entre dos participantes en un ambiente virtual. El experimento consiste en pares de participantes adultos con la tarea de explorar un ambiente virtual (un espacio unidimensional) con un ratón de computadora y recibir, como retroalimentación, estímulos táctiles en el dedo índice de la otra mano. El cursor de cada uno de los participantes en el ambiente virtual corresponde a un campo receptor de 4 unidades de extensión. Cuando el cursor (o avatar, usando términos del ambiente virtual) de un participante cruza un obstáculo en el ambiente, recibe la estimulación en el dedo, en caso contrario no existe tal estimulación. La figura 5.1 muestra gráficamente la disposición física de los participantes.

Dentro del ambiente virtual, además del otro avatar, cada uno de los individuos puede encontrarse con otros dos objetos: uno de ellos es un “objeto sombra” de un individuo el cual replica sus movimientos; y el otro objeto es estático al cual permanece en un punto fijo en el entorno (ver Fig. 5.2). Los participantes permanecen vendados y la tarea es explorar el ambiente dentro del cual cada uno puede percibir un total de tres objetos: el ‘avatar’ (campo receptor del otro participante), un objeto sombra y un objeto estático. El objetivo de la tarea es dar click cuando las personas perciban que están interactuando con el otro participante y solamente en esa situación. Lo anterior implica que el objetivo no es solamente identificar objetos en movimiento, sino identificar aquel objeto que sienta y responda al encuentro del otro participante.

Al finalizar el experimento y después del correspondiente análisis, se llevó a cabo una clasificación de los eventos que pudieron determinar la respuesta de los participantes así como sus estrategias. Primero, la frecuencia de estimulación debida al avatar, al objeto sombra o al objeto estático, siendo los resultados: 52.2 % para el avatar, seguido del objeto estático con 32.7 % y finalizando con el objeto sombra con 15.2 %. La segunda

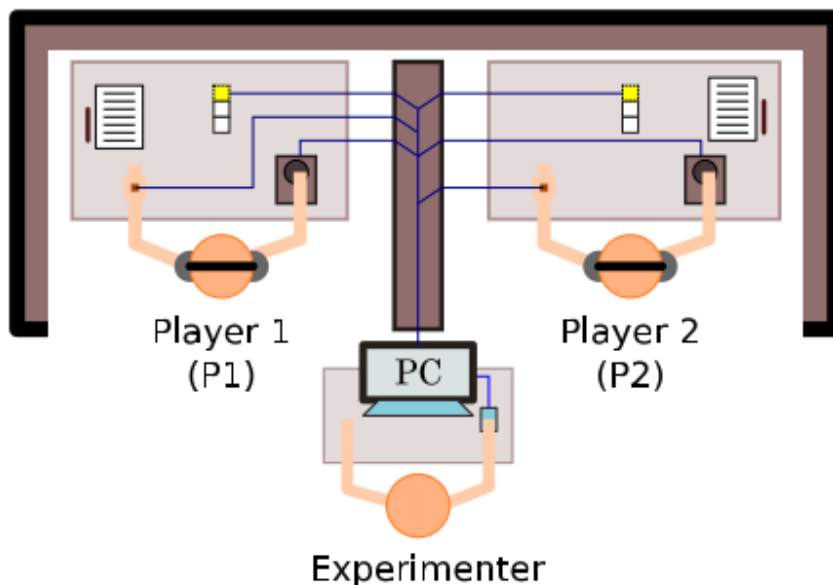


Figura 5.1: Diseño del experimento del paradigma del cruce perceptual. La interacción que se da entre los participantes es por medio de un ambiente virtual (una interfaz humano-computadora). Cada participante tiene a su alcance un ratón que permite el control de su “avatar” en el ambiente virtual y también cuentan con un dispositivo portátil de retroalimentación háptica que vibra cada vez que el avatar se traslapa con algún objeto del ambiente. (Imagen tomada de Froese et al. 2014)

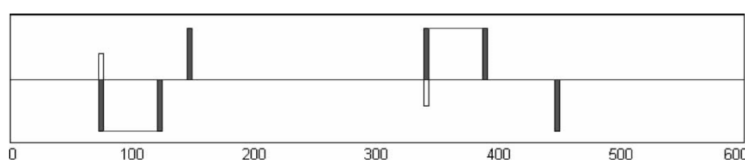


Figura 5.2: Diseño del ambiente virtual. Representación del ambiente virtual donde los participantes tienen la tarea de identificar cuál, de los objetos que están en el ambiente, es el otro participante. Los avatares de los participantes son los que tienen los campos receptores (recuadros blancos). Los objetos sombra se encuentran a una distancia fija de los avatares (48 unidades). Y finalmente los objetos estáticos se encuentran en posiciones fijas en 148 y 448 unidades. (Imagen tomada de Froese & Di Paolo 2010)

estadística fue el cálculo de la causa del número de clicks arrojando los siguientes

resultados: 65.9 % para el avatar, 23.0 % para el objeto sombra y 11.0 % para el objeto estático. Los resultados obtenidos son evidencia que la distinción que hay sobre el otro participante y el resto de los objetos surge principalmente de la dinámica de interacción, además de ser más eficiente que la capacidad individual de distinguir si la entidad con la que se está interactuando es tal o cual objeto.

Otro de los hechos evidentes fue que a través de una interacción oscilatoria sobre los objetos encontrados es que se da en mayor medida una interacción con el otro individuo. Así mismo, es posible determinar que la situación más estable de interacción es entre los individuos y no con los otros objetos. Lo anterior se debe a que la búsqueda de ambos agentes permite que estén más tiempo juntos, además de que el tipo de interacción entre ellos es distinta con respecto a la que se da con respecto a los otros dos objetos. Por ejemplo, en relación con el objeto estático no hay una respuesta de sensibilidad ante la presencia de un individuo y eventualmente se rompe la interacción que existe entre ellos. Con respecto al objeto sombra es distinto, ya que se puede percibir movimiento, pero solamente el individuo es el que puede recibir la estimulación o retroalimentación; mientras que el otro individuo -el del objeto sombra- no recibe estimulación y sigue su trayecto exploratorio, lo que provoca que eventualmente se rompa el proceso de interacción.

Por otro lado, investigadores en el área de la robótica evolutiva consideraron interesante la generación de un modelo artificial que permitiera el estudio de este experimento desde una óptica diferente. En la siguiente sección se hace un recorrido por los distintos modelos sintéticos para el análisis de la cognición.

5.2.1. Modelos sintéticos mínimos de cognición social

En la década de los 90s la robótica evolutiva se comenzó a utilizar en la creación de modelos sintéticos para el estudio de la cognición, lo que más tarde se conocería como "modelos mínimos de comportamiento cognitivo" (cf. Beer, 1996, 2003; Harvey et al., 2005).

Los modelos mínimos de comportamiento cognitivo permiten reducir en gran medida el número de variables del fenómeno, lo que abre la posibilidad de tener un mayor control de una gran parte de las características y así poder determinar el tipo de comportamiento a analizar. Con esto, es posible el estudio de la cognición en un nivel de complejidad reducido, así la posibilidad de tener una amplia comprensión de todo el sistema (Froese et al., 2012). Una de las partes centrales de la robótica evolutiva es la implementación de algoritmos evolutivos, los cuales dan la oportunidad de explorar de diferentes formas el análisis de un experimento y al mismo tiempo no agregan un gran sesgo en el diseño de los componentes del modelo. De tal manera que se pueden crear modelos con la capacidad de cubrir e integrar estrategias novedosas. Además, la robótica evolutiva integra el enfoque dinámico, del que ya se habló anteriormente, como parte importante para estudiar la cognición y también cómo las interacciones de un sistema pueden generar comportamientos que no podrían ser obtenidos si se tuvieran módulos aislados con comportamientos específicos.

Dentro del contexto de este marco de investigación se generó un creciente interés por utilizar los modelos sintéticos para el estudio de las dinámicas de interacción de la cognición social (e.g. Di Paolo, 2000; Quinn, 2001; Iizuka & Ikegami, 2004, 2007; Iizuka & Di Paolo, 2007; Di Paolo et al., 2008; Froese & Di Paolo, 2008).

5.2.2. Modelos sintéticos de cruce perceptual

Dentro de esta misma línea de investigación Ezequiel Di Paolo, Marieke Rohde y Hiroyuki Iizuka (2008) crearon el modelo sintético del experimento de cruce perceptual hecho por Auvray y sus colegas. En general el modelo posee las mismas características que el experimento original. Los controladores de los agentes artificiales están basados en CTRNNs. Algunas particularidades del proceso de creación del modelo fueron que la CTRNN que determina el comportamiento de los agentes fue evolucionada con un algoritmo genético que maximiza el desempeño de la tarea: localizar el otro agente y pasar el mayor tiempo posible tan cerca como sea posible, mientras evitan quedar atrapados con los otros objetos del ambiente. La dificultad con el procedimiento anterior fue que el algoritmo de búsqueda evolutivo no tenía la posibilidad de encontrar una solución satisfactoria. Los resultados fueron que los agentes se quedaban pegados a cualquier objeto que encontraran en el ambiente y permanecían ahí hasta el término de la prueba. Por ejemplo, si un agente se encontraba con otro agente el valor de aptitud obtenido era perfecto, lo que se presentaba en la mayoría de los casos. Sin embargo, el comportamiento anterior, no es, ni muy inteligente ni adaptativo, ya que no ejemplifica las estrategias desplegadas por los seres humanos, los cuales tienen periodos de exploración en el ambiente. Lo anterior se pudo solucionar agregando un retraso entre el cruce con un objeto y la sensación recibida por el agente. Con esta modificación las trayectorias de los agentes comenzaron a parecerse más a las de los participantes humanos. Sin embargo, surgía una pregunta como consecuencia de lo anterior: ¿las oscilaciones exploratorias de los agentes están relacionadas con el retraso de respuesta que existe en los humanos? Con el fin de responder cuestiones como la anterior y otras relacionadas con el tiempo se comenzaron a desarrollar modificaciones del modelo sintético de Di Paolo y sus colegas.

Para continuar con esta línea de investigación Froese y Di Paolo (2010) hicieron algunas modificaciones al modelo que replicó el experimento original de cruce perceptual (Di Paolo et al., 2008). Las modificaciones muestran que se pueden evolucionar agentes lo suficientemente robustos para seguir resolviendo la tarea de interactuar juntos, a pesar de que se cambien algunas características del ambiente como se describe a continuación. La primer modificación que se hizo fue intercambiar los campos receptores de los agentes, es decir, lo que sentaba el agente A era recibido por el agente B y viceversa. La segunda modificación fue probar cómo reaccionaban los agentes ante el conflicto de intereses: los agentes fueron evolucionados para que, en vez de permanecer lo más cerca posible del otro agente, estuvieran lo más cerca posible del objeto sombra. En ambos casos se pudo completar la tarea a pesar de los cambios hechos al modelo, lo que muestra que a pesar de los objetivos individuales, las interacciones del sistema determinan

-en cierta medida- el comportamiento de los individuos, aumentando y extendiendo, en ciertas situaciones, las capacidades de los individuos involucrados.

Los resultados obtenidos muestran que, bajo ciertas circunstancias, el objetivo global puede ser alcanzado aún anteponiéndose a los objetivos individuales. Además se sustenta la idea de que las capacidades individuales y aquéllas que emergen de las interacciones y la dinámica del sistema se complementan una a la otra. Lo anterior indica que el hecho de resolver la tarea designada no es un logro puramente individual. Por otro lado, una herramienta que aportó gran información fue el análisis dinámico, convirtiéndose en una estrategia novedosa en el estudio de los experimentos. El análisis mostró que el sistema no es meramente reactivo ante estimulaciones externas, más bien que el comportamiento de los agentes depende de los estados anteriores. Diferentes estados de la CTRNN imitan una memoria a corto plazo que modifica y da características muy particulares al comportamiento del sistema.

Con este tipo de modelos, la robótica evolutiva ofrece la oportunidad de plantear nuevas hipótesis para generar nuevos experimentos que contribuyan a aprobar o refutar esas hipótesis, como en el trabajo realizado por Froese y Di Paolo (2011). En dicho trabajo fue implementado un modelo de robótica evolutiva donde se pudieron observar diferentes estrategias de los agentes para resolver la misma tarea. La organización del ambiente fue diseñada con dos importantes modificaciones: evolucionar los agentes en condiciones ambiguas. La primer modificación fue hacer el tamaño de los objetos lo suficientemente pequeño, y la segunda los objetos sombra se encuentran a una distancia lejana de sus respectivos avatares. Estas modificaciones generaron comportamientos más complejos y se pudieron observar otro tipo de estrategias basadas en la dinámica de las interacciones con los demás objetos. En el segundo caso, se pudo observar que es más complicado detectar la sensibilidad que se puede percibir del otro agente que usar la información del ambiente.

A pesar de que se han desarrollado diferentes modelos sintéticos inspirados en el paradigma de cruce perceptual, ninguno implementó el reconocimiento explícito por medio de clicks. En esta tesis se analiza un modelo que muestra la relevancia de la interacción relacionada con la percepción mutua y su reconocimiento explícita. Como ya fue mencionado, los trabajos anteriores desarrollaron las simulaciones utilizando la cercanía relativa entre los agentes como señal de “reconocimiento” del otro individuo. Sin embargo, a pesar de que los agentes permanezcan mucho tiempo juntos, eso no implica un reconocimiento explícito como tal entre ellos. Por tanto esta sería la primera ocasión en que un modelo de robótica evolutiva del experimento de cruce perceptual implementa este aspecto crucial para los estudios psicológicos.

Modelo de robótica evolutiva para el reconocimiento explícito de agencialidad

Esta sección describe las características del modelo de simulación de robótica evolutiva, así como también de los detalles del algoritmo evolutivo y de la red neuronal recurrente de tiempo continuo empleados para generar el comportamiento de los agentes. En la sección 5.1 se hace una descripción del diseño del ambiente virtual, los objetos, las características físicas de los agentes y de sus controladores. En la sección 5.2 se explica el procedimiento que se siguió para poder implementar de manera exitosa el modelo de robótica evolutiva, descripción de las etapas evolutivas, así como de la función de aptitud para evolucionar a los agentes. En la sección 5.3 se muestran los resultados obtenidos del modelo, ejemplos representativos de algunas ejecuciones y gráficas del comportamiento de los agentes. Para la sección 5.4 se hace un análisis del comportamiento de los agentes, de las estrategias utilizadas, así como una comparación con respecto a otros trabajos anteriormente desarrollados. Por último, en la sección 5.5 se emplea un paquete de sistemas dinámicos para analizar la dinámica de las interacciones del sistema.

6.1. Diseño

El modelo de simulación incluye dos agentes que, siguiendo el trabajo de Auvray et al. (2009), están frente a frente en un ambiente unidimensional (ver Fig. 6.1). El ambiente unidimensional es una línea que está unida por los extremos y que tiene una extensión de 600 unidades de espacio, es decir, el ambiente describe una trayectoria circular con una circunferencia de 600 unidades. Las capacidades de los agentes son moverse sobre la línea horizontal en ambas direcciones, izquierda y derecha, además de poder dar “click” (posteriormente se hará una descripción más detallada de esta característica). Los agentes miden 4 unidades de longitud y poseen un campo receptor, el cual detecta la presencia de un objeto en el ambiente, generando un 1 si detecta un

6. MODELO DE ROBÓTICA EVOLUTIVA PARA EL RECONOCIMIENTO EXPLÍCITO DE AGENCIALIDAD

objeto y 0 en caso contrario. Además de los agentes con las características anteriormente descritas, también hay otros objetos dentro del ambiente: objetos sombra y objetos estáticos. Los objetos sombra son objetos que replican el movimiento de los agentes, ya que están fijos a cierta distancia de ellos (150 unidades). Los objetos estáticos están fijos en 148 y 448 unidades de distancia para cada lado.

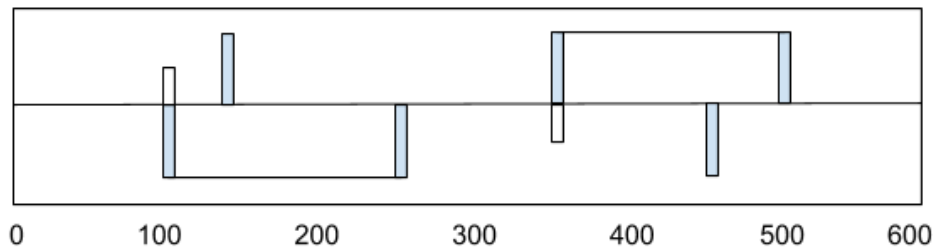


Figura 6.1: Diseño del ambiente virtual. Dos participantes pueden moverse a través del ambiente virtual que consta de una línea que se une por sus extremos y que tiene una longitud de 600 unidades. Los avatares son los únicos objetos del ambiente que cuentan con un campo receptor (rectángulo blanco), el cual es capaz de detectar tres tipos de objetos en el ambiente: el otro avatar, un objeto sombra y un objeto estático. El objeto sombra correspondiente a cada uno de los avatares está localizado a una distancia fija de ellos (150 unidades). Los objetos estáticos se encuentran a 148 y 448 unidades con respecto al comienzo del ambiente, cada uno es visto por su correspondiente avatar. Todos los objetos del ambiente tienen las mismas dimensiones.

El comportamiento de cada uno de los agentes es determinado por lo que simula su sistema nervioso: una red neuronal recurrente de tiempo continuo. La CTRNN modela la dinámica interna de los controladores de cada agente. Una misma CTRNN es utilizada para evolucionar el comportamiento de ambos agentes. Las variables independientes son los parámetros de la CTRNN que se describirán a detalle más adelante y las variables dependientes son el número de estimulaciones y el número de clicks.

La CTRNN consiste de 3 nodos totalmente conectados con los demás nodos y con sí mismos. Todos los nodos de la CTRNN reciben el valor de entrada que es el valor del sensor multiplicado por una ganancia r_i . Cada uno de los agentes tiene dos motores, izquierdo y derecho, que le permiten desplazarse por el ambiente. La velocidad de los motores está dada por el valor de salida de las neuronas 1 y 2, que se encuentra dentro del rango $[0, 1]$. Por lo tanto, la velocidad total del agente es el resultado de la resta de la salida de la neurona 1 menos el valor de salida de la neurona 2. La salida de la neurona 3 determina el comportamiento del click. Para poder simular el comportamiento del click en los agentes artificiales se implementó una restricción física (como la que se puede

observar en los humanos). Primero, fue necesario convertir el valor continuo de la salida de la neurona en un valor binario de 0 o 1. Para eso se estableció que si la salida de la tercer neurona es mayor a 0.75 eso indica que se apretó el botón de click, en caso contrario, cuando la salida de la neurona es menor a 0.75, el botón no está presionado. La simulación de restricciones físicas fue que el click debe de durar como mínimo 50 pasos de tiempo y como máximo 300. Además, se estableció un límite máximo para el número total de clicks por agente, el cual fue de 40 clicks.

Un genoma del algoritmo genético contiene los parámetros de la CTRNN: las constantes de tiempo, los sesgos, los pesos de las conexiones de las neuronas, las ganancias de entrada, y las ganancias de salida. El algoritmo genético evoluciona los valores de estos parámetros que fueron establecidos en los dentro de los siguientes rangos: constantes de tiempo [1, 200], sesgos [-3, 3], pesos entre las neuronas [-8, 8], ganancias del sensor [-100, 100], ganancias de los motores [-5, 5]. Se agregó un retraso en la activación del sensor de 50 pasos de tiempo, debido a que sin este retraso, no se pudieron obtener comportamientos robustos de los agentes. Como ya se había mencionado en la sección anterior, en trabajos previos como el de Di Paolo et al. (2008), se utilizó también un retraso para poder generar comportamientos robustos y adaptativos. Una probable justificación de este retraso es que el ser humano no tiene una capacidad de reacción instantánea frente a un estímulo, como sí lo pueden tener los agentes artificiales. Para la mutación en el algoritmo genético se utilizó una distribución de probabilidad gaussiana con una varianza de 0.02, la cual es aplicada sobre el vector de búsqueda.

6.2. Procedimiento

La tarea a realizar para cada uno de los agentes es interactuar con su entorno durante 800 unidades de tiempo y que la mayor parte de esa interacción sea con el otro agente mientras identifican explícitamente esa interacción dando clicks. La única información que reciben los agentes es la obtenida por el sensor. Para lograr que los agentes pudieran resolver la tarea, fueron necesarias dos etapas de evolución. La primera etapa los agentes fueron evolucionados en un ambiente sin el resto de los objetos (objetos sombra y estáticos), con el objetivo de que desarrollaran el comportamiento del click. La función de aptitud para esta primera etapa fue la siguiente:

$$F = \left(\frac{1}{T} \sum_0^T 1 - \frac{d(t)}{300} \right) \times 0.7 + \#Clicks \times 0.3 \quad (6.1)$$

Donde $d(t)$ es la distancia entre los agentes en el tiempo t . La parte de la distancia entre los agentes representa el 70 % de la función de aptitud, mientras que el número de clicks el 30 %. El número de generaciones fue de 2000 generaciones y el tamaño de la población de 100 individuos. En la segunda etapa de la evolución se añadieron los objetos sombra y los objetos estáticos, y se mantuvieron fijos el resto de los parámetros de la primera fase, así como la función de aptitud. En cada generación los agentes fueron evaluados 36 veces, cada una de estas evaluaciones asignaba posiciones iniciales

6. MODELO DE ROBÓTICA EVOLUTIVA PARA EL RECONOCIMIENTO EXPLÍCITO DE AGENCIALIDAD

distintas a los agentes. Por sí mismo, el algoritmo evolutivo, colocaría como mejores soluciones cuando los agentes se encuentren en posiciones de ventaja, es decir, como cuando los agentes están juntos al inicio de la evaluación, porque de esa manera obtendrán un mejor valor de aptitud. Por otro lado, estaría asignando poco valor de aptitud a aquellas evaluaciones en las que los agentes se encuentren con los objetos estáticos y se queden interactuando con ellos sin continuar explorando el ambiente. Sin embargo, el comportamiento esperado es que independientemente de las posiciones iniciales los agentes sean capaces de resolver la tarea, es decir, explorar el ambiente y a través de esa exploración, encontrar el otro agente y reconocerlo por medio del click. Por tanto, lo anterior hace que el algoritmo evolutivo no tenga un buen desempeño en determinadas circunstancias.

Para evitar lo anterior, se recalculó la aportación que hacen cada una de las evaluaciones al valor de aptitud. El procedimiento fue realizar un ordenamiento de mayor a menor con respecto al valor de aptitud de las evaluaciones y su contribución fue inversamente proporcional esa clasificación. Este recálculo asegura que las posiciones iniciales que son difíciles de evolucionar y que obtuvieron bajos valores de aptitud tengan mayor peso que las sencillas. Cada una de las evaluaciones tuvo un paso de tiempo de 0.1, dando un total de 8000 pasos de tiempo. Con respecto a la CTRNN a evolucionar, al inicio de cada evaluación el estado interno de las neuronas, el valor de entrada y el de salida fueron iniciados con 0. Se agregó ruido al movimiento de los motores de los agentes, el valor fue tomado de una distribución uniformemente distribuida entre un rango de $[-0.125, 0.125]$. Se agregó ruido a los motores debido a que los agentes son controlados por la misma CTRNN, lo que implica que haya movimientos espejo impidiendo romper la simetría cuando, por ejemplo, ambos agentes están interactuando con los objetos estáticos al mismo tiempo. El ruido agregado permitió generar soluciones más robustas.

6.3. Resultados

Los primeros intentos para que los agentes pudieran resolver la tarea no resultaron ser satisfactorios. Primero se comenzó con la idea de replicar los modelos de cruce perceptual de trabajos anteriores usando como función de aptitud la distancia entre los agentes y posteriormente agregar el comportamiento del click. Sin embargo, dado que el algoritmo genético se encarga de maximizar la función de aptitud, los resultados no correspondían con el objetivo establecido, ya que los agentes se quedaban interactuando solamente con el objeto estático o con el objeto sombra. Lo anterior básicamente dependía de las posiciones iniciales de las ejecuciones y no llevaba a resultados lo suficientemente robustos y adaptativos a las diferentes configuraciones y situaciones del sistema. Con el objetivo de evitar esos resultados se agregó un retraso de 50 pasos de tiempo entre cruzarse con un objeto en el ambiente y el registro del sensor. Esa modificación cambió el comportamiento y la estrategia de interacción entre los agentes, originando movimientos oscilatorios cuando se encontraban con un objeto.

La figura 6.2 muestra el comportamiento de los agentes que se identifican mutuamente por medio de clicks a lo largo de toda una ejecución. Se puede observar cómo va cambiando la señal del “click” con respecto a las interacciones. Por otro lado y como ya se mencionaba anteriormente, el hecho de agregar el retraso al recibir la señal de retroalimentación por medio del sensor fue algo que le agregó robustez al sistema. Sin embargo, no fue posible obtener el comportamiento del reconocimiento explícito con clicks incluyendo todos los objetos en el ambiente desde el inicio del proceso evolutivo. Razón por la cual se tuvo que separar en dos fases la evolución de los agentes.

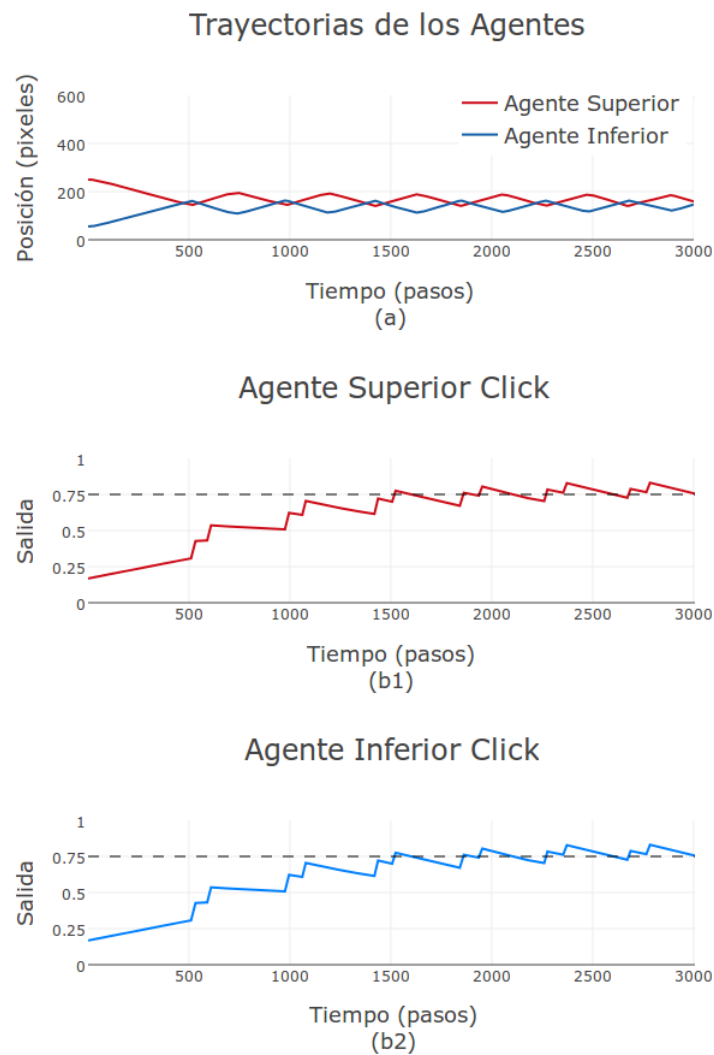


Figura 6.2: Ejemplo del comportamiento de los agentes. La figura (a) muestra cómo es el proceso de interacción entre los agentes. En el siguiente par de figuras (b1 y b2) se muestra el comportamiento del click para ambos agentes.

6. MODELO DE ROBÓTICA EVOLUTIVA PARA EL RECONOCIMIENTO EXPLÍCITO DE AGENCIALIDAD

Como en los modelos anteriores de robótica evolutiva se puede observar un comportamiento oscilatorio en los agentes, lo que permite en determinadas situaciones romper la interacción con los objetos de los cuales no se tiene una correspondencia a las estimulaciones recibidas (objeto sombra y estático). Además se puede observar que después de algunas interacciones el comportamiento del click se hace presente en la dinámica del sistema, lo que muestra que el sistema no es totalmente reactivo a las señales externas, sino hasta que realmente existe un proceso de interacción es cuando surge el reconocimiento explícito.



Figura 6.3: Ilustración del comportamiento de los agentes durante una ejecución de ejemplo. En la figura a) se observa la trayectoria de los agentes y su interacción con los diferentes objetos. En las figuras b1) y b2) se pueden observar la salida de las neuronas que controlan los motores izquierdo y derecho de los agentes.

En la figura 6.3 se integran el resto de los objetos en el ambiente y se observa la nueva dinámica que toma el sistema. En la trayectoria de los agentes se puede identificar: primero, el agente inferior comienza interactuando con el objeto estático, mientras que el agente superior interactúa con el objeto sombra del otro agente; segundo, se rompen las correspondientes interacciones que tiene cada agente y terminan finalmente interactuando entre los agentes mismos. Además de las trayectorias de los agentes, se observan las salidas de los motores izquierdo y derecho que determinan el movimiento de los agentes. Para las neuronas que controlan los motores existe una regulación entre los valores que aporta una y otra a la velocidad del agente. Para determinar el movimiento de los agentes en el ambiente se podía haber asignado solamente una neurona para tal tarea. Sin embargo, es más sencillo que se genere una regulación y adaptabilidad a la dinámica si son dos neuronas las que están involucradas.

Para la parte del comportamiento del click en la figura 6.4 se muestra la salida de la neurona que controla este evento de los agentes. Como se puede apreciar en la figura 6.2 el comportamiento del click no es reactivo a los eventos que suceden en el entorno, sino que depende de una serie de interacciones con los elementos de ese entorno para que se haga o no click. Ese comportamiento obtenido es debido al tipo de red neuronal utilizada, ya que las acciones dependen de los estados previos del sistema.

6.3.1. Resultados estadísticos

Como en el experimento original del cruce perceptual, se realizaron los cálculos de la frecuencia de las estimulaciones y la causa de los clicks. La frecuencia de las estimulaciones fue clasificada en tres secciones debido al objeto que las originó: el avatar, el objeto sombra y el objeto estático. El promedio total de estimulaciones fue de 25.06. La mayor parte de las estimulaciones recibidas por los agentes fueron debidas justamente al otro agente con 56.94 % con desviación estándar de ± 19.96 , seguido del objeto estático con 29.58 % con desviación estándar ± 14.65 y al último el objeto sombra con 13.47 % con desviación estándar ± 12.23 .

Para contabilizar la causa de los clicks se tuvieron que tomar varias consideraciones, principalmente cuando se traslapan los objetos. Para determinar un “contacto” lo que se hizo fue tomar bloques continuos de estimulaciones y en el caso del traslapamiento se le dio prioridad al agente sobre el resto de los objetos. Como existe un retraso en el sensor, se tomó una ventana de 50 pasos de tiempo para conocer el último objeto que generó la estimulación. Como resultado del proceso anterior, se obtuvo que la mayoría de los clicks fueron debido al otro agente con 60.01 % y una desviación estándar de 22.84, luego el objeto estático con 28.22 % y desviación estándar de 15.85, finalizando con el objeto sombra con 11.76 % y 7.67 de desviación estándar. Estos cálculos muestran que el resultado de la evolución del comportamiento de los agentes enfatiza la coordinación mutua en vez de sólo el comportamiento del click por sí mismo.

El cálculo de la causa de los clicks muestra que los agentes hicieron click más frecuentemente en el avatar del otro agente que en el resto de los objetos del ambiente. Lo cual cumple exitosamente con el objetivo establecido para el modelo. En secciones

6. MODELO DE ROBÓTICA EVOLUTIVA PARA EL RECONOCIMIENTO EXPLÍCITO DE AGENCIALIDAD

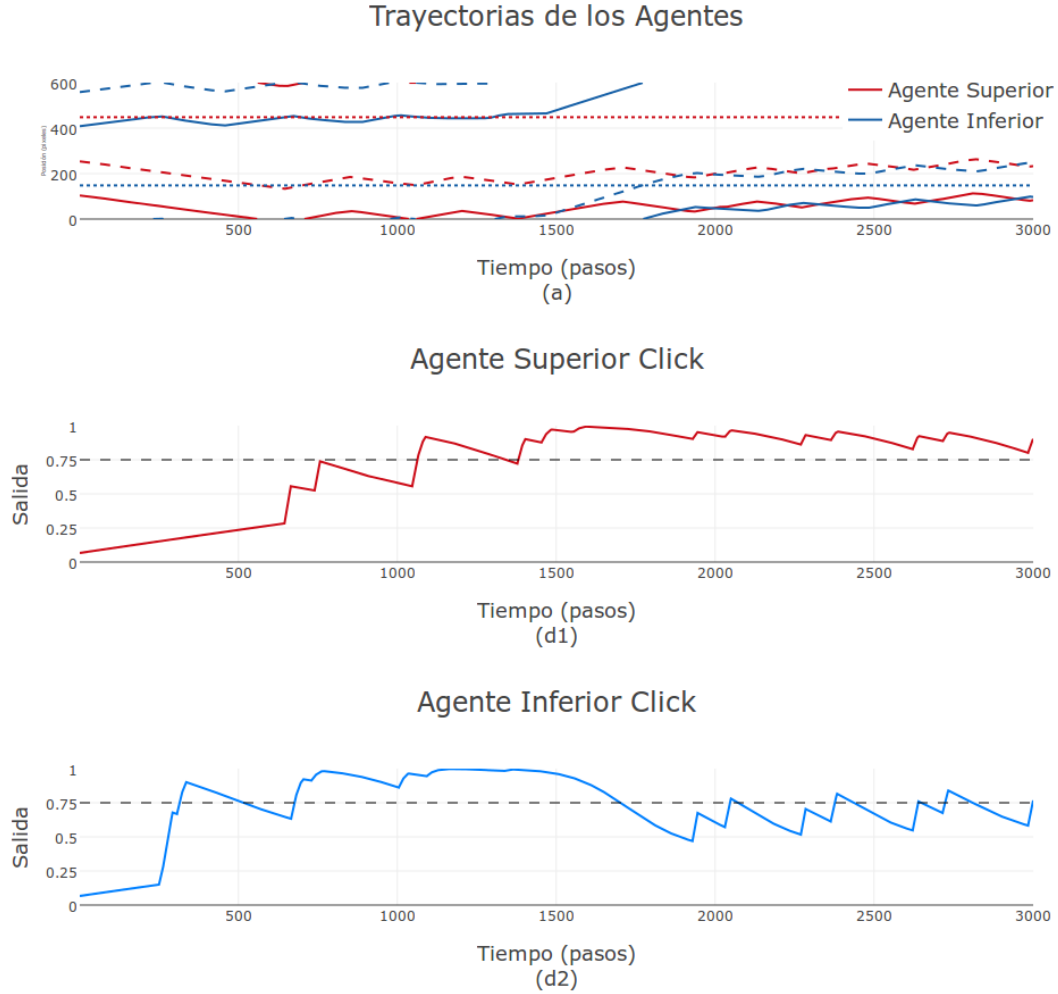


Figura 6.4: En esta serie de figuras se han incluido todos los objetos en el ambiente. En la figura (a) se muestran las trayectorias de los ambientes y su interacción con diferentes objetos. Se puede notar que al comienzo de las interacciones el agente inferior está interactuando con el objeto estático, pero eventualmente se rompe esa interacción y continúa explorando el ambiente hasta que se encuentra con el otro agente y permanecen juntos hasta el final de la corrida. Mientras tanto, el agente superior interactúa con el objeto sombra del agente inferior. (b1 y b2) Diferentes estimulaciones que reciben los agentes a lo largo de la corrida. Comportamiento de los motores del agente superior (c1) y agente inferior (c2). Comportamiento del click del agente superior (d1) y del agente inferior (d2).

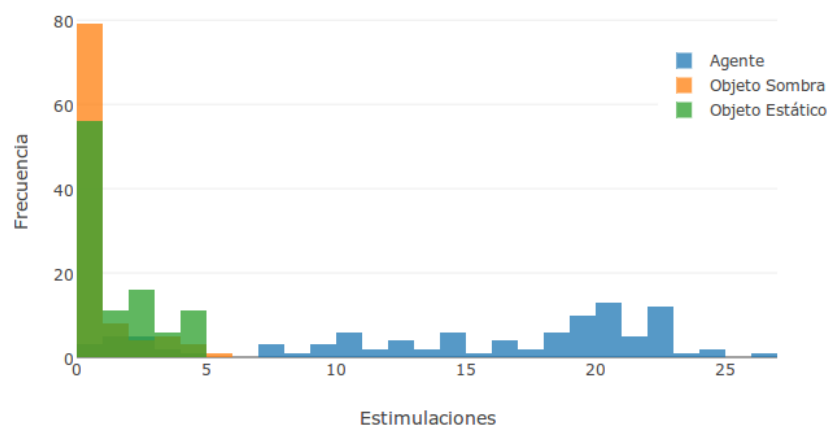


Figura 6.5: Histograma que muestra la frecuencia de las estimulaciones sobre cada uno de los objetos del ambiente: agente, objeto sombra y objeto estático.

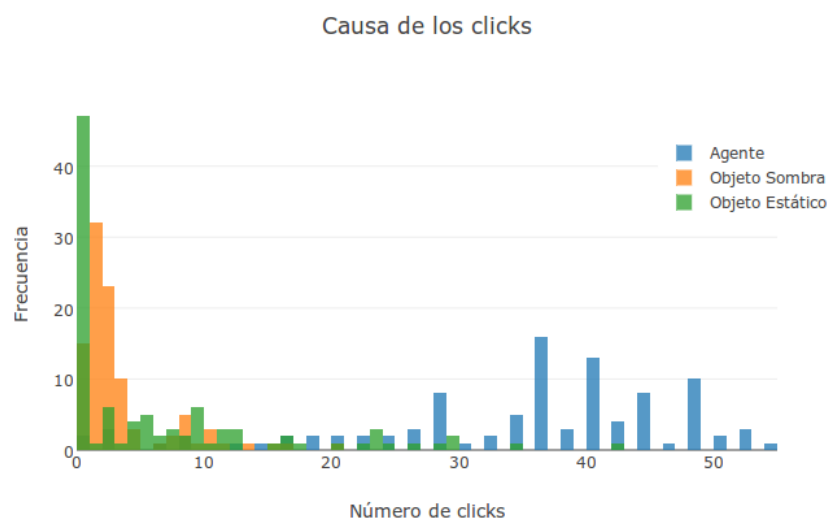


Figura 6.6: Histograma que muestra la cantidad de clics que fueron causados por los diferentes objetos. En ambos histogramas se puede apreciar el agrupamiento de los datos dependiendo del tipo de objeto.

6. MODELO DE ROBÓTICA EVOLUTIVA PARA EL RECONOCIMIENTO EXPLÍCITO DE AGENCIALIDAD

posteriores se podrán analizar y discutir con más detalles estos resultados, junto con el comportamiento desplegado por los agentes.

Las figuras 6.5 y 6.6 son histogramas que representan de la frecuencia de las estimulaciones y de la causa de los clicks, respectivamente. Para obtener estas estadísticas, fueron ejecutadas 100 instancias del programa desde diferentes posiciones iniciales. En estas gráficas existe una marcada tendencia en el número de estimulaciones y número de clicks entre los agentes. Es decir, para los agentes se presentan frecuencias altas y para los objetos sombra y estático frecuencias bajas.

Por otro lado y finalmente después de distintos intentos con diferente número de nodos, la estructura de la CTRNN resultante fue la siguiente:

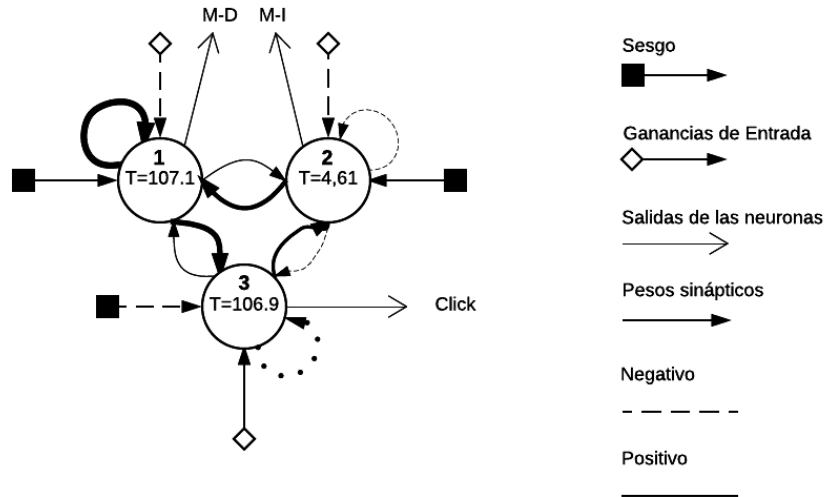


Figura 6.7: Estructura de la CTRNN de los agentes que mejor pudieron resolver la tarea.

Los círculos representan los nodos de la red neuronal y las constantes de tiempo τ_i , las flechas que comienzan con un cuadrado son los sesgos de las conexiones b_i , las que comienzan con un diamante son las ganancias de entrada r_i , las flechas normales representan los pesos de las conexiones entre los nodos w_{ij} . Las conexiones con valores negativos fueron representadas con línea punteada, mientras que las conexiones positivas con líneas sólidas. El grosor de las líneas es aproximadamente proporcional al valor de los parámetros.

6.4. Análisis

Una de las motivaciones por las cuales es de interés el modelo del cruce perceptual es para conocer las diferentes estrategias generadas por los agentes para resolver exitosamente la tarea. Esta cuestión será analizada en esta sección con el uso de sistemas dinámicos para entender la dinámica que pertenece al sistema.

Como se puede observar en las gráficas mostradas anteriormente, los agentes describen movimientos oscilatorios sobre los objetos que se encuentran en el ambiente virtual. Ésto eventualmente permite que exista la discriminación entre el otro agente y el resto de los objetos (objeto sombra y estático). Tanto el experimento de Auvray y sus colegas como los modelos de robótica evolutiva muestran que la dinámica del proceso interactivo de todo el sistema es lo que genera un comportamiento ‘social’ adecuado para poder resolver la tarea establecida. A continuación se despliegan gráficas de los diferentes estados que adquiere el sistema dinámico de reconocimiento explícito de agencialidad con respecto al tiempo.

Particularmente en este modelo mínimo de cognición social se puede observar que en las trayectorias del sistema se genera un punto estable (ver Fig. 6.8). No obstante, que las trayectorias converjan hacia un punto eso no significa que exista solamente un comportamiento, ya que en el modelo además de que los agentes se encuentran cerca pueden ejecutar el comportamiento del click.

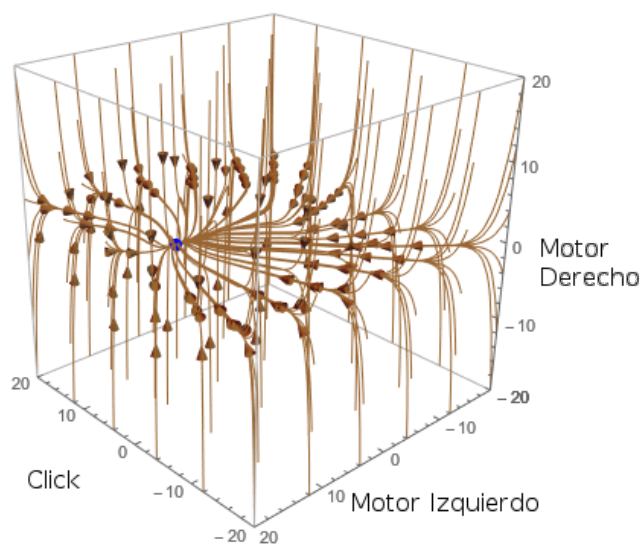


Figura 6.8: Trayectorias de la CTRNN. Flujo de trayectorias representativas que muestran una perspectiva general de la naturaleza dinámica del sistema. El punto azul indica el punto estable del sistema dinámico.

6. MODELO DE ROBÓTICA EVOLUTIVA PARA EL RECONOCIMIENTO EXPLÍCITO DE AGENCIALIDAD

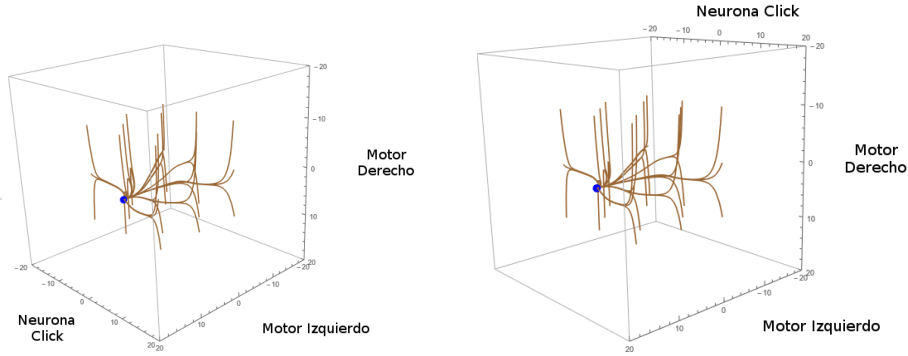


Figura 6.9: Perspectiva global de los atractores. Los ejes representan las salidas de sigmoides de los tres nodos de la CTRNN. En la gráfica de la izquierda el valor de los sensores es 0 con un punto estable en (11.73, 1.71, 2.20), mientras que para el valor de entrada 1 el atractor fue localizado en (12.90, 2.99, 2.54).

Es decir, los agentes pudieron ser evolucionados con la capacidad de desarrollar dos comportamientos. Por otro lado, es importante hacer notar que las figuras de esta sección muestran las trayectorias de las salidas de las neuronas de manera aislada, no en el proceso de interacción. Debido a ésto, en realidad no sólo existe un punto como atractor del sistema, sino que existen dos dependiendo de cual sea el valor asignado al parámetro de entrada I que es el que recibe la información de los sensores: 0 ó 1. La gráfica 6.9 muestra los atractores para el caso en que el valor de los sensores sea 0 y 1. De esta forma, el sistema de los controladores se encuentra oscilando entre estos dos atractores para poder ejecutar eficientemente su comportamiento con el fin de reconocer al otro agente. Sin embargo, estas gráficas, sólo muestran los eventos de manera aislada y no en el proceso de interacción con el entorno. Si los “cerebros” artificiales son analizados de manera aislada describen trayectorias que caen en un sólo punto de equilibrio fijo a la vez, como ya se vio en las gráficas 6.9. Pero, cuando se monitorea la actividad neuronal correspondiente a la oscilación del comportamiento entre los dos atractores, podemos observar que se generan complicados ciclos dinámicos adquiriendo una forma de *histéresis*. La explicación para este fenómeno es que el análisis no tiene que estar enfocado en un sólo cerebro, sino debe ser adoptado un enfoque más amplio que considere la dinámica de las interacciones como punto fundamental para explicar el éxito en el comportamiento observado.

6.5. Discusión

El modelo de robótica evolutiva reportado en este trabajo revela que dos agentes artificiales interactuando en un ambiente minimalista dan un número de clicks signi-

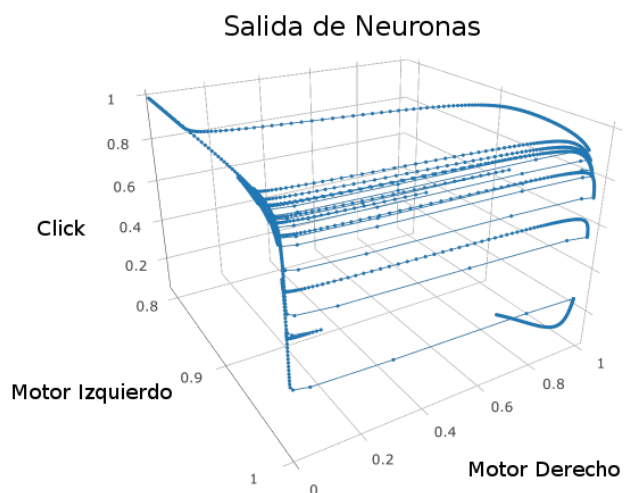


Figura 6.10: Salida de las tres neuronas: motor derecho, motor izquierdo y click.

ficativamente mayor entre ellos mismos que con respecto a los otros objetos. La alta proporción de respuestas correctas es consecuencia de los movimientos exploratorios hechos por los agentes. Dichos movimientos convergen en el surgimiento estrategias conjuntas en relación a la exploración mutua. Lo anterior puede ser explicado debido a la dinámica sensomotora de la interacción entre ellos, es decir, el algoritmo genético selecciona a los pares de individuos que están cerca y que además dan click cuando están interactuando con el otro agente. En términos generales, se ven favorecidas las situaciones en las que los agentes se encuentran juntos. Los resultados tanto estadísticos como del comportamiento demuestran que el rol de la percepción es crucial para poder completar la tarea de identificación del otro agente exitosamente. Además el proceso de percepción conjunta (cuando dos agentes se perciben mutuamente) es el resultado de la interacción entre los agentes y la discriminación dinámica del comportamiento con respecto al resto de los objetos.

Por otra parte, las contingencias sensomotoras son partes fundamentales para determinar el comportamiento, ya que se presentan oscilaciones simétricas y asimétricas dependiendo el objeto con el cual se esté interactuando. Este experimento refleja que aún cuando los agentes no son explícitamente conscientes del entorno, es decir, no saben cuáles son las características del ambiente virtual y tampoco pueden ver los movimientos de los otros objetos en el ambiente, aún bajo esas condiciones es posible hacer una contribución suficiente para poder cumplir con la tarea. El comportamiento mostrado es resultado del esfuerzo mutuo de cada uno de los agentes para poder encontrar al otro y reconocerlo usando el click.

Durante las diferentes simulaciones realizadas se pudieron identificar dos tipos de interacciones: las estables y las inestables. Abordando primero las interacciones ines-

6. MODELO DE ROBÓTICA EVOLUTIVA PARA EL RECONOCIMIENTO EXPLÍCITO DE AGENCIALIDAD

tables, en éstas se encuentran interactuando ya sea el agente con el objeto sombra del otro agente o con el correspondiente objeto estático. En dichos casos la estimulación sólo la recibe un agente y no existe una correspondencia a sus movimientos debido a que el otro objeto no tiene la capacidad de percibir al agente. No obstante, cuando se analiza el comportamiento de los agentes, se pueden observar interacciones estables como consecuencia de que ambos tienen una manera de percibir al otro. Adicionalmente, hay dos características importantes en el comportamiento de los agentes: no se encuentran en una posición fija y se encuentran muy presentes en la interacción. Esas dos características hacen que exista más probabilidad de que interactúen entre ellos que con los otros objetos.

En otro orden de ideas, una de las situaciones que tomó mayor trabajo evolucionar fue la interacción con el objeto estático debido a cómo fue escrita la función de aptitud. Tomó muchos intentos llegar a evolucionar el comportamiento que mostrara una discriminación entre los objetos. En los primeros intentos, los agentes se quedaban estancados interactuando con el objeto estático ya que de esa forma probablemente no obtenían un gran valor de aptitud, pero el que obtenían les era suficiente para poder quedar en una buena posición para ser elegidos en el proceso de selección del algoritmo evolutivo. Sin embargo y a pesar de que existe cierta estabilidad y un patrón en la interacción con el objeto estático, para que se de un reconocimiento entre los dos agentes es imprescindible cómo percibe cada agente los movimientos del otro objeto en relación a los suyos. Una parte de ese problema fue solucionado con el retraso entre el cruce con un objeto y el sensado, además de agregar ruido a las salidas de los motores.

Así mismo y de acuerdo con Di Paolo et al. (2008), es importante hacer énfasis en que no fue necesario evolucionar a los agentes con un módulo dedicado a interacción social o especializado en reconocimiento de estrategias de movimientos. Simplemente la dinámica de las interacciones pudieron generar el comportamiento correcto para lograr el objetivo. Además, esa dinámica permite que se pueda resolver una tarea que parece imposible desde una perspectiva individual, ya que -como fue mencionado antes- los agentes no tienen manera de ver las trayectorias de los demás objetos.

Dos importantes contribuciones del presente trabajo fueron las siguientes: 1) se replicaron con valores bastante cercanos el alto porcentaje de clicks correctos obtenidos en Froese et al. (2014) y 2) se mostró que este alto porcentaje es debido probablemente a los procesos de interacción (ya que, como se presentó previamente, las dinámicas aisladas no tienen los suficientes elementos para explicar el comportamiento observado).

6.6. Conclusiones

En esta sección se presentó a detalle la implementación del modelo de robótica evolutiva integrando el reconocimiento explícito de agencialidad. Se describieron a detalle el diseño del modelo, el procedimiento para evolucionar el comportamiento de los agentes, los resultados obtenidos de las simulaciones y un análisis que contempló el uso de la teoría de sistemas dinámicos.

Haciendo uso de la robótica evolutiva se puso crear un modelo que revelara la importancia de la dinámica en el reconocimiento de otra entidad con autonomía de entre otro tipo de objetos. Debido a la disposición física del ambiente y la única información que está disponible para los agentes, la dinámica del movimiento es determinante para poder concluir la tarea exitosamente. Si bien se podría presentar un caso hipotético en el que un agente no se mueva, sólo de esa forma el otro agente no tendría oportunidad de distinguir entre este agente y el objeto estático respectivo. Ésto muestra la importancia de la dinámica y de la interacción entre agentes, y que en ausencia de estos elementos no se podría generar un comportamiento robusto y complejo como el mostrado.

Otra particularidad que se puede ver tanto en el experimento original como en el modelo presentado en este trabajo es que los individuos tienen que distinguir entre dos tipos de dinámicas distintas: la primera es la dinámica de interacción con todos los objetos del ambiente y la segunda es la dinámica de la percepción conjunta (Lenay & Stewart, 2012). En la primera dinámica quien tiene la facultad de percibir cualquier otro objeto solamente son los agentes, mientras que la segunda dinámica surge exclusivamente entre los agentes. La dinámica de percepción conjunta favorece e induce las situaciones en las que los individuos se encuentran muy cercanos entre sí, lo que a su vez propicia que aumente la probabilidad de recibir el estímulo y eventualmente dar click.

El análisis dinámico permitió mostrar de una manera más clara que enfocándose en los agentes de manera aislada no es posible deducir el comportamiento que se despliega en un ambiente colectivo. Sin embargo, cuando se analiza la dinámica colectiva del sistema es posible conocer los estados en los que oscilan los agentes. Ese comportamiento colectivo pareciera ser generado por agentes con capacidades cognitivas complejas, sin embargo todo es debido a las interacciones situadas en el ambiente.

Conclusiones

El objetivo principal de esta tesis fue alcanzado al poder generar un modelo de simulación del experimento original del cruce perceptual. Eso implicó poder implementar en el modelo el comportamiento del click, algo que no fue desarrollado por ninguno de los trabajos anteriores. Otro de los objetivos fue analizar y estudiar por medio de un modelo sintético mínimo de cognición social los mecanismos que subyacen a la actividad de identificar a otro sujeto con intencionalidad. El cruce de percepciones entre los sujetos o agentes juega un papel crucial para que se de ese reconocimiento.

Una cuestión que permitió poder tener un modelo con las características necesarias fue el haber implementado los controladores de los agentes utilizando CTRNN. La dinámica interna de los controladores de los agentes permite que su comportamiento dependa de no solamente de las circunstancias inmediatas, sino también en la historia que se genera de su interacción con el ambiente y los elementos que componen a éste. Además, los individuos pueden o no pueden ser conscientes de las capacidades colectivas antes de que surjan, lo que abre las puertas para poder explotar y hacer uso de esta característica de los sistemas dinámicos.

No se puede seguir viendo las características externas como cuestiones inertes e independientes a agentes artificiales o a los individuos. Parte del comportamiento de los individuos está determinado por la manera en que se da la interacción con el ambiente, además de las capacidades físicas que cada uno de ellos posee. La combinación de los elementos anteriores es lo que hace que emerja un comportamiento complejo a partir de agentes que por sí mismos no tienen capacidades cognitivas complejas.

Sin embargo, las pruebas de concepto, las conclusiones a las que se han llegado y las aplicaciones de estas investigaciones se deben en gran medida a las herramientas producidas en el área de la inteligencia artificial. Es importante resaltar que la utilización de modelos sintéticos ha permitido poder explorar diferentes aspectos que de repente parecerían imposibles o por lo menos muy difíciles de replicar en la realidad. El uso de modelos de simulación de fenómenos naturales facilita diferentes cosas, por ejemplo, es posible reducir la dimensionalidad y complejidad de un problema para su correspondiente análisis. También del análisis de los patrones de las simulaciones es posible explorar las relaciones con respecto a los correspondientes términos teóricos

7. CONCLUSIONES

análogos a los patrones que se encuentran en la naturaleza.

Para construir el modelo aquí presentado no fue necesario el uso de los conceptos tradicionales de la inteligencia artificial clásica como representación del conocimiento, crear módulos específicos de interacción, jerarquización de acciones, etc. No fue necesaria la especificación explícita de las distintas acciones que debía de ejecutar el agente ni de condicionar su comportamiento dependiendo de la situación en la que se encontrara. Si ese hubiera sido implementado no se hubiera generado un comportamiento lo suficientemente robusto y complejo, o probablemente hubiera consumido muchos recursos para poder lograrlo.

Además de estudiar la cognición por sí misma, los avances e investigaciones en esta área han podido impactar a otras áreas y generar proyectos muy interesantes. Por ejemplo, Marti (2010) tomó como referencia el paradigma del cruce perceptual para desarrollar dispositivos interactivos capaces de tener una regulación mutua en actividades realizadas en conjunto. También se han desarrollado robots para interactuar con distintas discapacidades.

Una parte del análisis central que fue presentado en el capítulo anterior se relaciona mucho con el tema de las personas con autismo. Este trastorno es caracterizado por la alteración en la interacción social y el comportamiento restringido y repetitivo. Las personas con este padecimiento necesitan prestar muchísima atención a las señales externas para poder hacer algo y obtener la aprobación social de tal acción, en vez de responder directamente a la presencia de la otra persona. Relacionado con lo presentado en este trabajo, el modelo mostró que fue mucho más complicado evolucionar a los agentes para tener una capacidad sensitiva a la respuesta del otro agente.

Algunas de las cosas que se tienen que enfrentar y que no se pueden omitir cuando se estudia la cognición social es integrar la dinámica colectiva del sistema. En la mayoría de los casos se considera que el estudio de la dinámica y las interacciones de un sistema es algo complejo, sin embargo con las herramientas correctas probablemente esa complejidad sea mucho menor que si se estudiaran los elementos por separado. Por otra parte, en la cognición social ha sido complicado determinar cuáles son las características que son aportadas por el individuo y cuáles surgen de la dinámica como tal de todo el sistema. Como se mencionó en los primeros capítulos, la investigación de la cognición se ha visto marcada por el estudio de los fenómenos de forma aislada y lejanos al *tiempo real*. Estos aspectos se consideran cruciales para el estudio de la cognición a diferentes escalas y niveles. Así mismo, los resultados mostrados aquí pueden incrementar la relación entre el área de robótica evolutiva y la ciencia cognitiva.

Para finalizar, con el desarrollo del enactivismo la idea es que el estudio de la cognición sea lo más cercano a lo que percibimos que sucede en la naturaleza. Además de hacer un replanteo en diferentes aspectos, desde conceptual, en cuanto a las herramientas, a los paradigmas, a los cuestionamientos, a los resultados. Algo que sin duda no sólo marcaría una revolución para la inteligencia artificial, sino y debido a que el enactivismo es un enfoque interdisciplinario, ese impacto sería en muchas más áreas.

7.1. Trabajo a futuro

Como parte de la implementación de modelos artificiales de simulaciones siempre han existido diversos retos a contemplar en su implementación. Específicamente para este trabajo la cuestión del diseño de la función de aptitud fue un aspecto muy complicado de implementar. Lo anterior fue debido a la naturaleza del fenómeno a analizar, además de respetar la propiedad de hacer pocas suposiciones y restricciones en el diseño. Una de las propuestas que resultaron bastante prometedoras para implementarse es la de dividir el comportamiento en diferentes fases, de tal manera se comiencen con comportamientos simples y gradualmente se vayan generando soluciones más complejas, construidas a través de las sencillas (Lehman & Stanley, 2011). Otra opción interesante es hacer una variación de ambientes para que el proceso de evolución sea mucho más rápido que cuando no existe tal variación (Kashtan et al., 2007).

Con respecto a los modelos de cruce perceptual una implementación interesante sería replicar uno de los experimentos hechos por Lenay & Stewart (2012), donde es el mismo diseño que el experimento original pero la retroalimentación es por medio de sonidos. Sin duda arrojaría nueva información interesante en la investigación de la cognición social.

Bibliografía

- Ashby, W. R. (1952). *Design for a Brain*. London: Chapman and Hall.
- Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Chapman and Hall LTD.
- Auvray, M., Lenay, C., & Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, 27(1), pp. 32–47.
- Bar-Yam, Y. (2002). General Features of Complex Systems. In *Encyclopedia of Life Support Systems*. Oxford, UK.
- Barandiaran, X., & Moreno, A. (2006). On what makes certain dynamical systems cognitive: A minimally cognitive organization program. *Adaptive Behavior*, 14(2).
- Beer, R. (2003). The Dynamics of Active Categorical Perception in an Evolved Model Agent. *Adaptive Behavior*, 11, pp. 209–243.
- Beer, R. D. (1995). A dynamical systems perspective on agent–environment interaction. In *Artificial Intelligence*, 72. Essex, UK: Elsevier Science Publishers Ltd., pp. 173–215. doi: 10.1016/0004-3702(94)00005-L
- Beer, R. D. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behavior. In P. Maes, M. J. Matarić, J.-A. Meyer, J. Pollack, & S. W. Wilson (Eds.) *From Animals to Animats 4: Proc. of the 4 th Int. Conf. on Simulation of Adaptive Behavior*. Cambridge, MA: The MIT Press, pp. 421–429.
- Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3), pp. 91–99.
- Beer, R. D., & Gallagher, J. C. (1992). Evolving Dynamical Neural Networks for Adaptive Behavior. *Adaptive Behavior*, 1(91).
- Brooks, R. A. (1991). Intelligence without representation. In *Artificial Intelligence*, 47(1-3). pp. 139–159. doi: 10.1016/0004-3702(91)90053-M

BIBLIOGRAFÍA

- Brooks, R. A. (1997). From earwigs to humans. *Robotics and Autonomous Systems*, 20(2), pp. 291 – 304. doi: 10.1016/S0921-8890(96)00064-4
- Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, 9(1), pp. 55–72.
- Chiel, H., & Beer, R. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neurosciences*, 20(12), pp. 553–557.
- Clark, A. (1997). *Being There: Putting brain, body, and world together again*. Cambridge, MA: The MIT Press.
- Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York: Oxford University Press.
- D’Andrade, R. (2003). *The Development of Cognitive Anthropology*. Cambridge.: Cambridge: Cambridge University Press.
- De Jaegher, H., Di Paolo, E., & Adolphs, R. (2016). What does the interactive brain hypothesis mean for social neuroscience? A dialogue. In *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 371(1693). The Royal Society. doi: 10.1098/rstb.2015.0379
- De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14, pp. 441–447.
- Di Paolo, E. (2000). Behavioral coordination, structural congruence and entrainment in a simulation of acoustically coupled agents. *Adaptive Behavior*, 8(1), pp. 25–46.
- Di Paolo, E., & De Jaegher, H. (2012). The interactive brain hypothesis. *Frontiers in Human Neuroscience*, 6(163). doi: 10.3389/fnhum.2012.00163
- Di Paolo, E., Rohde, M., & De Jaegher, H. (2007). Horizons for the enactive mind: Values, social interaction, and play. In *Cognitive Science Research Paper*, 587. University of Sussex, Brighton, UK.
- Di Paolo, E. A., Rohde, M., & Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New Ideas in Psychology*, 26(2), pp. 278–294.
- Dreyfus, H. L. (1972). *What computers can’t do: A critique of Artificial Reason*. Harper and Row.
- Dreyfus, H. L. (1992). *What computers still can’t do*. Cambridge: MIT Press.

- Elman, J. (1998). Connectionism, Artificial Life, and Dynamical Systems: New approaches to old questions. In W. Bechtel, & G. Graham (Eds.) *A Companion in Cognitive Science*. Basil Blackwood, Oxford.
- Eysenck, M. W. (1990). *Cognitive Psychology: An International Review..* West Sussex, England: John Willey & Sons.
- Fodor, J. (2000). *The Mind Doesn't Work That Way*. The MIT Press.
- Fraisse, P., Piaget, J., & Reuchlin, M. (1963). *Experimental psychology: Its scope and method. 1. History and method..* New York: Basic Books.
- Froese, T. (2010). From cybernetics to second-order cybernetics: A comparative analysis of their central ideas. *Constructivist Foundations*, 5(2), pp. 75–85.
- Froese, T., & Di Paolo, E. A. (2008). Stability of coordination requires mutuality of interaction in a model of embodied agents. In M. Asada, J. C. T. Hallam, J.-A. Meyer, & J. Tani (Eds.) *From Animals to Animats 10: Proceedings of the 10th International Conference on Simulation of Adaptive Behavior*. Berlin, Germany: Springer-Verlag, pp. 52–61.
- Froese, T., & Di Paolo, E. A. (2010). Modelling social interaction as perceptual crossing: an investigation into the dynamics of the interaction process. *Connection Science*, 22(1), pp. 43–68.
- Froese, T., & Di Paolo, E. A. (2011). Toward minimally social behavior: social psychology meets evolutionary robotics. In I. K. G. Kampis, & E. Szathmari (Eds.) *Advances in Artificial Life: Darwin Meets von Neumann. 10th European Conference, ECAL 2009*. Berlin, Germany: Springer Verlag, pp. 426–443.
- Froese, T., Iizuka, H., & Ikegami, T. (2014). Embodied social interaction constitutes social cognition in pairs of humans: A minimalist virtual reality experiment. In *Scientific Reports*, 4. doi: 10.1038/srep03672
- Froese, T., Lenay, C., & Ikegami, T. (2012). Imitation by social interaction? Analysis of a minimal agent-based model of the correspondence problem. *Frontiers in Human Neuroscience*, 6.
- Goldberg, D. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading, MA: Addison-Wesley.
- Graham, G. (2016). Behaviorism. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2016 ed. URL <http://plato.stanford.edu/archives/fall2016/entries/behaviorism/>.
- Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42, pp. 335–346.

BIBLIOGRAFÍA

- Harvey, I., Di Paolo, E. A., Wood, R., Quinn, M., & Tuci, E. A. (2005). Evolutionary Robotics: A new scientific tool for studying cognition. *Artificial Life*, 11(1-2), pp. 79–98.
- Hodges, A. (2012). *Alan Turing: The Enigma*. Princenton University Press.
- Holland, J. (1975). *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press.
- Iizuka, H., & Di Paolo, E. (2007). Minimal agency detection of embodied agents. In F. A. e Costa, L. Rocha, I. H. E. Costa, & A. Coutinho (Eds.) *Advances in Artificial Life: Proceedings of the 9th European Conference on Artificial Life*. Berlin, Germany: Springer-Verlag, pp. 485–494.
- Iizuka, H., & Ikegami, T. (2004). Adaptability and diversity in simulated turn-taking behavior. *Artificial Life*, 10(4), pp. 361–378.
- Iizuka, H., & Ikegami, T. (2007). Turn-taking interaction as a cooperative and co-creative process. *Infant Behavior and Development*, 30(2), pp. 278–288.
- Izquierdo, E., Harvey, I., & Beer, R. D. (2008). Associative learning on a continuum. *Adaptive Behavior*, 16(6), pp. 361–384.
- James, I. (2009). Claude Elwood Shannon 30 April 1916 — 24 February 2001. In *Biographical Memoirs of Fellows of the Royal Society*, 55. The Royal Society, pp. 257–265. URL <http://rsbm.royalsocietypublishing.org/content/55/257>. doi: 10.1098/rsbm.2009.0015
- Kashtan, N., Noor, E., & Alon, U. (2007). Varying environments can speed up evolution. In *Proceedings on the National Academy of Sciences of the USA*, 104(34). Cambridge, MA: The MIT Press, pp. 13711–13716.
- Lehman, J., & Stanley, K. (2011). Abandoning objectives: Evolution through the search for novelty alone. In *Evolutionary Computation*, 19(2). Cambridge, MA: The MIT Press, pp. 189–223.
- Lenay, C., & Stewart, J. (2012). Minimalist approach to perceptual interactions. *Frontiers in Human Neuroscience*, 6:98. doi: 10.3389/fnhum.2012.00098
- Lowe, C., & Rotenberg, A. (2012). Cognitive Science Hexagon. URL <https://commons.wikimedia.org/w/index.php?curid=19461955>.
- Malone, J. C. (2009). *Psychology: Pythagoras to Present*. Cambridge, MA.
- Marter, M., & Shieh, J. C. (2010). *Guide to Research Techniques in Neuroscience*. Amsterdam: Elsevier/Academic Press.
- Marti, P. (2010). Perceiving while being perceived. *International Journey of Design*, 4(2), pp. 27–38.

- Maturana, H. R., & Varela, F. J. (1987). *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston, MA: Shambhala Publications.
- McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence 4*, pp. 463–502.
- McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955). A proposal for the Dartmouth summer research project on Artificial Intelligence. Funding proposal. (First documented use of the term ‘Artificial Intelligence’). Retrieved from: <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>.
- McClelland, J., Rumelhart, D., & Hinton, G. (1986). The appeal of parallel distributed processing. In J. M. D. Rumelhart, & the PDP Research Group (Eds.) *Parallel Distributed Processing*, I. Cambridge, MA: MIT Press, pp. 3–40.
- McCorduck, P. (2004). *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. Natick, MA: A. K. Peters., 2nd ed.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), pp. 115–133. doi: 10.1007/BF02478259
- Newell, A., Shaw, J. C., & Simon, H. (1959). Report on general problem-solving program. In *Proceedings of the International Conference on Information Processing..* Paris, pp. 256–264.
- Newell, A., & Simon, H. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist*, 26(2), pp. 145–159.
- Nolfi, S., & Floreano, D. (2000). *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-organizing Machines*. Cambridge, MA: MIT Press.
- Nolfi, S., & Floreano, D. (2002). Synthesis of autonomous robots through artificial evolution. *Trends in Cognitive Sciences*, 6, pp. 31–37.
- O’Regan, J., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), pp. 939–973.
- Pfeifer, R., & Bongard, J. C. (2007). *How the Body Shapes the Way We Think: A New View of Intelligence (Bradford Books)*. The MIT Press.
- Quinn, M. (2001). Evolving Communication without Dedicated Communication Channels. In J. Kelemen, & P. Sosík (Eds.) *Advances in Artificial Life: 6th European Conference..* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 357–366. doi: 10.1007/3-540-44811-X_38
- Robinson, H. (2016). Dualism. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 ed. URL <http://plato.stanford.edu/archives/win2016/entries/dualism/>.

BIBLIOGRAFÍA

- Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, purpose and teleology. *Philosophy of Science*, 10(1), pp. 18–24.
- Russell, D. (1987/2004). *The Oxford Companion to the Mind*. Oxford University Press, 2nd ed.
- Shapiro, L. (2011). *Embodied Cognition*. New York: Routledge.
- Siciliano, B., & Khatib, O. (2008). *Springer Handbook of Robotics*.
- Silva, F., Correia, L., & Christensen, A. L. (2016). Evolutionary Robotics. 11(7), pp. 33333. revision #168217.
- Thagard, P. (2014). Cognitive Science. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2014 ed. URL <http://plato.stanford.edu/archives/fall2014/entries/cognitive-science/>.
- Thelen, E., & Smith, L. (1998). Dynamic systems theories. In *Handbook of child psychology*, 1. NJ: Wiley, pp. 258–312.
- Thompson, E., & Varela, F. J. (2001). Radical Embodiment: Neural Dynamics and Consciousness. *Trends in Cognitive Science*, 5(10), pp. 418–425.
- Turing, A. (1936). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42, pp. 230–265.
- van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, 21(5), pp. 615–665.
- van Gelder, T., & Port, R. F. (1995). It's About Time: An Overview of the Dynamical Approach to Cognition. In R. F. Port., & T. van Gelder (Eds.) *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: The MIT Press, pp. 1–43.
- Varela, F. J. (1979). *Principles of Biological Autonomy*. New York: Elsevier/North Holland.
- Varela, F. J. (1991). Organism: A meshwork of selfless selves. In A. I. Tauber (Ed.) *Organisms and the Origins of Self*. Dordrecht, Netherlands: Kluwer Academic Publishers, pp. 79–107.
- Varela, F. J. (1997). Patterns of Life: Intertwining Identity and Cognition. In *Brain and Cognition*, 34(1). pp. 72–87.
- Varela, F. J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: The MIT Press.

- von Foerster, H. (1979). *Cybernetics of cybernetics*. New York: Gordon and Breach, krippendorff k. ed.
- Ward, D., & Stapleton, M. (2012). Es are good : Cognition as enacted , embodied , embedded , affective and extended. In F. Paglieri (Ed.) *Consciousness in Interaction: The Role of the Natural and Social Context in Shaping Consciousness*. Amsterdam: John Benjamins, pp. 89–105.
- Wiener, N. (1948). *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge, MA: The MIT Press.