



# VNiVERSIDAD D SALAMANCA

---

CAMPUS DE EXCELENCIA INTERNACIONAL

FACULTAD DE CIENCIAS

DEPARTAMENTO DE INFORMÁTICA Y AUTOMÁTICA

TESIS DOCTORAL:

## PRINCIPIOS MATEMÁTICOS DEL COMPORTAMIENTO NATURAL

Depositada por Sergio Miguel Tomé para obtener  
el título de Doctor en la Universidad de Salamanca

---

Dirigida por:

Dr. Ángel Luis Sánchez Lázaro y Dr. Luis Alonso Romero

Salamanca, Julio de 2017



---

El **Dr. Ángel Luis Sánchez Lázaro**, profesor titular del Departamento de Informática y Automática del área de Lenguajes y Sistemas Informáticos de la Universidad de Salamanca y el **Dr. Luis Alonso Romero**, Catedrático Jubilado del área de Ciencia de la Computación e Inteligencia Artificial de la Universidad de Salamanca

HACEMOS CONSTAR:

Que el trabajo de investigación que se recoge en la presente tesis doctoral, titulada **Principios Matemáticos del Comportamiento Natural** por **D. Sergio Miguel Tomé** para optar al Grado de Doctor con mención de Internacionalidad, ha sido realizado bajo nuestra dirección en el Departamento de Informática de la Universidad de Salamanca y tiene nuestro visto bueno.

Salamanca, 1 de Julio de 2017

Dr. Ángel Luis Sánchez Lázaro  
Profesor Titular  
Universidad de Salamanca

Dr. Luis Alonso Romero  
Catedrático Jubilado  
Universidad de Salamanca



# Índice general

<b>Summary</b>	<b>xv</b>
0.1. Motivation, hypotheses, and goals . . . . .	xv
0.2. The General Theory of Exobehaviors . . . . .	xvi
0.3. The Multifunctional Robots On Topological Notions Program . . . . .	xvii
0.4. Results and conclusions . . . . .	xvii
<b>Resumen</b>	<b>xix</b>
0.5. Motivación, hipótesis y metas . . . . .	xix
0.6. La teoría general del exocomportamiento . . . . .	xx
0.7. El programa robots multifuncionales mediante nociones topológicas . . . . .	xxi
0.8. Resultados y conclusiones . . . . .	xxi
<b>Agradecimientos</b>	<b>xxiii</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	2
1.2. Hipótesis de trabajo . . . . .	3
1.2.1. Hipótesis fundamentales . . . . .	3
1.2.2. Hipótesis tecnológicas . . . . .	4
1.3. Objetivos de la investigación . . . . .	5
1.4. Evolución de la investigación . . . . .	6
1.5. Esquema general . . . . .	7
<b>2. Teorías sobre los comportamientos</b>	<b>11</b>
2.1. Explicaciones de los comportamientos hasta el siglo XIX . . . . .	11
2.2. Investigaciones de los comportamientos durante la primera mitad el siglo XX . . . . .	13
2.3. Explicaciones de la inteligencia a partir de la segunda mitad del siglo XX . . . . .	15
2.4. Descubrimientos sobre los comportamientos de distintos organismos . . . . .	22
2.4.1. Cognición animal . . . . .	23
2.4.2. Las plantas y sus comportamientos . . . . .	25

2.4.3.	Los hongos y sus comportamientos . . . . .	28
2.4.4.	Los microorganismos y sus comportamientos . . . . .	28
2.4.5.	Comportamientos de apuesta . . . . .	29
2.5.	Inteligencia: Un concepto confuso e incapaz de unificar . . . . .	29
2.6.	Alternativas a la inteligencia . . . . .	32
2.7.	Conclusiones del análisis . . . . .	33
<b>3.</b>	<b>Evolución y comportamientos</b>	<b>35</b>
3.1.	Herencia y desarrollo en organismos biológicos . . . . .	36
3.2.	La formulación general de la selección natural . . . . .	38
3.3.	Las bases genéticas del comportamiento . . . . .	39
3.4.	La plasticidad fenotípica . . . . .	40
3.5.	Evolución y comportamiento . . . . .	41
3.6.	Conclusiones de la revisión . . . . .	42
<b>4.</b>	<b>Teoría general del exocomportamiento</b>	<b>45</b>
4.1.	La explicación científica . . . . .	46
4.2.	Razonando sobre sistemas naturales . . . . .	46
4.2.1.	Las reglas de razonamiento de Newton . . . . .	47
4.2.2.	Un método para razonar sobre sistemas físicos . . . . .	48
4.3.	Exocomportamientos: el conjunto de fenómenos . . . . .	49
4.4.	Fasa: causa única de los exocomportamientos . . . . .	55
4.4.1.	Análisis matemático de la fasa . . . . .	58
4.5.	Identificación experimental de tipos de fasa fundamental . . . . .	61
4.6.	Leyes de la fasa y predicciones para sistemas exoactivos . . . . .	66
4.7.	Reinterpretación de investigaciones previas sobre comportamientos . . . . .	68
4.8.	Análisis de la investigación sobre la TGE . . . . .	69
4.9.	Novedades de la reformulación . . . . .	73
<b>5.</b>	<b>La TGE y su conexión con la física</b>	<b>75</b>
5.1.	Orígenes del concepto de computación y su difusión en la física . . . . .	76
5.2.	Orígenes del concepto de información y su propagación en la física . . . . .	85
5.3.	Argumentos en contra de que el universo sea una computadora . . . . .	87
5.3.1.	El contraejemplo del sistema radioactivo . . . . .	88
5.3.2.	El contraargumento del continuo . . . . .	88
5.3.3.	El contraejemplo de la evaluación . . . . .	89
5.3.4.	El argumento de las infinitas dimensiones . . . . .	90
5.3.5.	El argumento del esquema lagrangiano . . . . .	91
5.3.6.	El contraejemplo de los oráculos . . . . .	92
5.4.	Propiedades fundamentales de la naturaleza y la TGE . . . . .	94
5.5.	Análisis de los resultados de la revisión . . . . .	95

<b>6. La fasa y el sistema nervioso I</b>	<b>97</b>
6.1. La descripción multinivel . . . . .	100
6.2. La actividad neuronal y la codificación de estados . . . . .	102
6.3. Lenguajes formales, redes neuronales y autómatas finitos deterministas	103
6.4. Un nuevo método de codificación neuronal: TLE . . . . .	104
6.5. Análisis de los resultados . . . . .	107
<b>7. La fasa y el sistema nervioso II</b>	<b>109</b>
7.1. Comparando sistemas nerviosos . . . . .	111
7.2. Sistemas nerviosos, predicción y la selección natural . . . . .	115
7.2.1. Revisando comportamientos, sistemas nerviosos y organismos .	122
7.2.2. Confrontando la hipótesis de la libertad de movimiento . . . . .	124
7.2.3. Alternativas a la hipótesis de la libertad de movimiento . . . . .	127
7.3. Características de los sistemas nerviosos . . . . .	128
7.3.1. Características biológicas del sistema nervioso . . . . .	129
7.3.2. Características computacionales del cerebro neuronal . . . . .	130
7.4. Sistemas de procesamiento neuronal frente a otros tipos de sistemas de procesamiento . . . . .	131
7.4.1. Los otros posibles sistemas de procesamiento . . . . .	131
7.4.2. Selección natural del sistema de procesamiento neuronal . . . . .	135
7.4.3. Principios del procesamiento neuronal . . . . .	140
7.5. Análisis de los resultados . . . . .	143
<b>8. La TGE y el paradigma computacional</b>	<b>147</b>
8.1. Revisión de las propuestas para el paradigma computacional . . . . .	150
8.1.1. Jerarquía de niveles . . . . .	150
8.1.2. Dominios . . . . .	153
8.2. Fundamentos matemáticos . . . . .	156
8.2.1. Semántica . . . . .	156
8.2.2. Teoría de la categoría . . . . .	159
8.3. Limitación de la propuesta de Mira y Delgado . . . . .	162
8.4. Descripción de la semántica- <i>t</i> de sistemas computacionales . . . . .	167
8.4.1. La semántica- <i>t</i> de los dominios de entrada . . . . .	168
8.4.2. La semántica- <i>t</i> de las acciones de un sistema . . . . .	171
8.4.3. El observador y el sistema de referencia . . . . .	173
8.5. Análisis de los resultados . . . . .	178
<b>9. TGE Y TCCV</b>	<b>185</b>
9.1. La teoría cognitiva de condiciones de verdad . . . . .	186
9.1.1. Estructuras de pasado presente futuro multivariadas con multi- opción . . . . .	187

9.1.2.	Jerarquía de lenguajes . . . . .	193
9.1.3.	Gramática reconocedora de condiciones de verdad . . . . .	195
9.2.	Mecanismos para realizar decisiones de la TCCV . . . . .	197
9.2.1.	Heurísticas de semántica cualitativa . . . . .	198
9.2.2.	Diferencias entre la TCCV y otros enfoques simbólicos . . . . .	199
<b>10.</b>	<b>Programa de investigación RMMNT</b>	<b>201</b>
10.1.	Representación de espacios para la navegación . . . . .	204
10.1.1.	Métodos cuantitativos . . . . .	204
10.1.2.	Métodos Cualitativos . . . . .	205
10.1.3.	Navegación basada en relaciones espaciales . . . . .	207
10.1.4.	El programa robots multifuncionales mediante nociones topológicas	208
10.1.5.	Representación mediante Espacios Topológicos . . . . .	208
10.2.	La heurística de semántica cualitativa topológica . . . . .	211
10.2.1.	El problema de navegación topológica cualitativa . . . . .	215
10.2.2.	Generalización a n-dimensiones . . . . .	215
10.3.	Extensiones de la HSCT: relaciones borrosas y de cercado . . . . .	217
10.3.1.	Extensión de HSCT añadiendo relaciones de cercado . . . . .	218
10.3.2.	Extension de la HCST mediante modificadores borrosos . . . . .	220
10.4.	Arquitectura de navegación cualitativa topológica . . . . .	224
10.4.1.	Fundamentos de la arquitectura de navegación cualitativa to- pológica . . . . .	225
10.4.2.	La ANCT 1.0, su implementación y prueba . . . . .	235
10.5.	Razonamiento cualitativo sobre orientación y giros . . . . .	243
10.5.1.	Premisas iniciales . . . . .	244
10.5.2.	Transformando información direccional en información topológica	244
10.5.3.	Heurística de semántica cualitativa direccional . . . . .	247
10.5.4.	La elección del grafo de razonamiento topológico direccional . . .	253
10.5.5.	Pruebas . . . . .	255
10.5.6.	Sistema de coordenadas cualitativas esféricas . . . . .	257
10.6.	Generación de lenguaje natural . . . . .	257
10.6.1.	Planteamientos para el lenguaje natural . . . . .	258
10.6.2.	El problema del anclaje del símbolo . . . . .	259
10.6.3.	Objetivos y planificación de la nueva etapa . . . . .	260
10.6.4.	Generando lenguaje natural mediante condiciones de verdad . . .	262
10.7.	Análisis de los resultados . . . . .	268
<b>11.</b>	<b>Conclusiones</b>	<b>273</b>
11.1.	Conclusiones sobre la TGE . . . . .	276
11.2.	Conclusiones sobre la TCCV y el programa RMMNT . . . . .	278
11.3.	Líneas futuras de investigación . . . . .	278

<b>12. Conclusions</b>	<b>281</b>
12.1. Conclusions about the GTE . . . . .	283
12.2. Conclusions about the TCCV and the MROTN program . . . . .	285
12.3. Future Research Directions . . . . .	285
<b>A. Axiomas de la TCCV</b>	<b>287</b>
<b>B. Alfabetos de metainformación</b>	<b>293</b>
<b>C. Código de la implementación de la ANCT 1.0</b>	<b>297</b>
<b>D. Topología</b>	<b>301</b>
<b>E. Conferencias y publicaciones realizadas</b>	<b>305</b>
<b>Bibliografía</b>	<b>307</b>



# Índice de figuras

1.1. Relaciones entre los capítulos . . . . .	9
2.1. Clasificación de propuestas para la inteligencia . . . . .	18
6.1. Niveles de la jerarquía de Newell . . . . .	100
6.2. Esquema del mecanismo Trayectorias-Lenguajes-Estados . . . . .	106
7.1. Clasificación del sistema nervioso . . . . .	116
8.1. Jerarquía de niveles de Marr . . . . .	151
8.2. Jerarquía de niveles de Pylyshyn . . . . .	151
8.4. Propuesta de niveles y dominios de Mira y Delgado . . . . .	154
9.1. Representación gráfica de una estructura PPFMM . . . . .	189
9.2. Representación gráfica de una perspectiva temporal . . . . .	191
10.1. Arquitecturas para robots . . . . .	203
10.2. Elemento de $\mathcal{B}_X \times \mathcal{B}_Y$ . . . . .	217
10.3. Ejemplos de situaciones de relaciones de cercado . . . . .	218
10.4. Relación topológica relativa y su complementaria . . . . .	221
10.5. Métodos de generación de puertas . . . . .	227
10.6. Diagrama de la ANCT 1.0 . . . . .	231
10.7. Ejemplo de pasos de la tarea <i>explorar</i> . . . . .	234
10.8. Ejemplo de pasos de la tarea <i>posicionarse para comer</i> . . . . .	236
10.9. Diagrama del motor de inferencia. . . . .	239
10.10. Proyección de un agente y un objeto . . . . .	246
10.11. Un agente girando con coordenadas egocéntricas y allocéntricas. . . . .	247
10.12. Pérdida de información direccional en la proyección. . . . .	248
10.13. Regiones de las variantes posicionales . . . . .	249
10.14. Proyección sin pérdida de información . . . . .	250
10.15. Grafo de razonamiento topológico direccional $GRDT_1^=$ . . . . .	251



# Índice de tablas

10.1. Casos para la asignación de un GLRT . . . . .	212
10.2. Tabla de relaciones topológicas relativas . . . . .	214
10.3. Funciones de verdad borrosas. . . . .	222
10.4. Tabla de medios y fines . . . . .	223
10.5. Representación de las acciones de $o_A$ . . . . .	241
10.6. Tabla con las funciones etiquetadas. . . . .	241
10.7. GGRTD que pueden existir entre un objeto y un agente. . . . .	251
10.8. Tabla de distancias y direcciones. . . . .	252
10.9. Tabla de direcciones y acciones. . . . .	253
10.10 Pares clave . . . . .	255
10.11 EPLP de núcleos . . . . .	266
10.12 Tabla del segundo nivel de la ELP . . . . .	266



# Summary

This doctoral thesis is centered around the search for a scientific theory that identifies biological and nonbiological autonomous behaviors. It also addresses the development of algorithms that allow for the creation of multifunctional robots. The work is focused on solving two problems. The first is to identify a physicalist theory capable of explaining how behaviors are produced. The second is to develop qualitative computational methods that contribute to the development of multifunctional robots. This second issue stems from human beings using and communicating with qualitative information in their everyday lives.

## 0.1. Motivation, hypotheses, and goals

Newton's mechanics, cell theory, the kinetic theory of gases, Maxwell's equations, and general relativity are examples of unification theories. Each of these brought about not only a new view of nature but also the development of new technologies and research. Unification theories are so beneficial to science and engineering that the effort required to search for a unification theory for any field of science is fully justified. Currently, the phenomena of behaviors are divided into disjointed sets and described individually, and many questions remain about behaviors. This, along with the importance of unification theories, led to the motivation for the work conducted in this doctoral thesis. The research that was carried out was based on the hypothesis of the General Theory of Exobehavior (GTE), which is capable of unifying behaviors. However, this theory is based on two hypotheses: the hypothesis of unification of behaviors and the hypothesis of the physicalism of behaviors. These hypotheses are outlined as follows.

*Hypothesis of the unification of behaviors:* All behaviors of biological and nonbiological systems are part of a single set of phenomena.

*Hypothesis of the physicalism of behaviors:* All behaviors that exist in reality can be explained in physical terms.

The General Theory of Exobehaviors purports that, given the two previous hypotheses,

all behaviors must be able to be explained only in physical terms. Another goal of this doctoral thesis was to develop computational methods that contribute to the creation of multifunctional robots capable of integrating into everyday human life. Thus, applied research requiring technological hypotheses was performed. Specifically, the hypotheses involved were the hypothesis of cognitivism based on true conditions and the hypothesis of multifunctional robots by topological notions. These two hypotheses make the following affirmations:

*Hypothesis of cognitivism based on true conditions:* True conditions are elements that permit the development of methods that provide all types of cognitive behaviors to robots and computers.

*Hypothesis of multifunctional robots by topological notions:* Qualitative methods for decision making and information communication that allow for the creation of multifunctional robots through the use of topological notions can be developed.

The combination of these two hypotheses guided the applied research to seek qualitative representations of space through topological notions.

## 0.2. The General Theory of Exobehaviors

The issue of finding a physicalist theory capable of unifying behaviors was addressed by developing and researching the GTE. This theory was developed by attempting to avoid issues observed in prior theories to explain behaviors and by possessing the positive characteristics previous proposals presented. Therefore, it is interesting to note that the GTE avoids mentalist concepts and is formulated from a physicalist standpoint. An unexpected consequence of the formulation is that it led to determining that ontogenic processes and behaviors belong to the same set of phenomena, the exobehaviors. A remarkable characteristic of the GTE is that its formulation presents a complete integration of exobehaviors with evolutionary processes. In addition to providing an explanation, the GTE has also served as a theoretical framework through which to raise questions about nature. Thus, questions about evolution have been raised that led to an investigation on computational models of multicellular organisms and the evolutionary forces that have shaped them. The GTE has shown two characteristics are needed for a theory of unification: explaining phenomena and directing new research.

### 0.3. The Multifunctional Robots On Topological Notions Program

The problem of developing qualitative computational methods that contribute to the creation of multifunctional robots has been addressed by proposing and developing the Multifunctional Robots on Topological Notions Program (MROTN). The objective of this research program is to develop qualitative methods using topological notions with which to define qualitative spatial relations. In addition to the use of topological notions, elements of the Cognitive Theory of True Conditions (CTTC) have been used. The program deals with the development of computational methods for functioning in settings and environments gradually. First, it attempts to find a method capable of facing the simplest condition possible. Once an effective method has been obtained, a more complex condition is posed, and the effectiveness of the method is examined. If it is not effective, the issues are investigated and a new method is sought that is effective and that complies with the conditions established by the program. Thus, this process is repeated continuously. It should be noted that the MROTN program is in agreement with the GTE, as the solutions found to contribute to the construction of computational robots fit with the properties of the exobehaviors exhibited by humans, and the objective of the program is to search for a particular solution to the general equation of the GTE.

### 0.4. Results and conclusions

Based on the research hypotheses, the objective of the thesis was to investigate and develop the GTE. That objective was based on two grounds. One reason was to have a framework to better understand behaviors and their effects on nature. The second reason was to guide the development of new methods to provide computers and robots with complex behaviors that allow them to integrate in wide-ranging human environments, facilitating inclusion in everyday life or helping in dangerous tasks. The fundamental research of this thesis produced various results. The literature review showed that intelligence and cognition are not concepts that can achieve unification as there are behaviors in nature that are not a product of information processing. Through the investigation and development of GTE, a more solid formulation was reached. One of the most significant discoveries was the finding that behaviors and ontogenic processes are part of the same set of natural phenomena since they have the same cause. In addition, the formulation led to the investigation of other issues from which significant results have been obtained. One is that computational robustness is an evolutionary advantage for the nervous systems of multicellular organisms. Another important result is that the semantics of model theory can be considered computational processes, and

that without them the computational process carried out by certain systems cannot be completely described. In addition, a new mechanism was proposed to interpret brain dynamics that would allow for the representation of categories. Regarding the applied research, the results showed that the creation of multifunctional robots through topological notions and true conditions are promising research directions. An architecture was developed, called qualitative topological navigation architecture 1.0, in which the decision making, objectives, and knowledge are stored in a qualitative way and allows for navigation in unknown and dynamic spaces. An interesting result obtained is a method for encoding directional information in topological relations. In addition, that result served to create a heuristic that permits decision making about spinning and defining a system of qualitative spherical coordinates that eliminates some limitations that the system of qualitative Cartesian coordinates has, which has been used since the beginning of the investigation. The last result obtained is a method for generating natural language expressions about qualitative spatial relations that opens the door to a new phase for research on the development of multifunctional robots based on topological notions and true conditions. The results obtained in this thesis lead to the conclusion that both the GTE and the MROTN program are solid research directions for obtaining results about behavior unification and the creation of multifunctional robots.

# Resumen

Esta tesis doctoral se enmarca dentro de la búsqueda de una teoría científica que unifique los comportamientos de sistemas biológicos y no biológicos autónomos. Además, también aborda el desarrollo de algoritmos que permitan la creación de robots multifuncionales. La tesis se centra en resolver dos problemas. El primero, encontrar una teoría fisicalista que sea capaz de explicar cómo se producen todos los comportamientos. El segundo problema es la necesidad de desarrollar métodos computacionales cualitativos que contribuyan a la creación de robots multifuncionales. Este segundo problema surge del hecho de que los seres humanos usan y se comunican información cualitativa en su vida cotidiana.

## 0.5. Motivación, hipótesis y metas

Ejemplos de teorías de unificación son la mecánica de Newton, la teoría celular, la teoría cinética de los gases, las leyes de Maxwell o la relatividad general. Cada una de ellas no ha provocado solo una visión nueva de la naturaleza, sino también el desarrollo de nuevas tecnologías e investigaciones. Los beneficios de las teorías de unificación para la ciencia y la ingeniería son tan altos que está plenamente justificado el esfuerzo en buscar una teoría de unificación para cualquier ámbito de la naturaleza. Actualmente los fenómenos de los comportamientos son divididos en múltiples conjuntos disjuntos y existen muchas preguntas sin respuesta sobre ellos. De ese hecho y la importancia que supone una teoría de unificación, emerge la motivación que ha conducido a realizar esta tesis doctoral. La investigación que se ha llevado a cabo está basada en la hipótesis de que la teoría general del exocomportamiento es una teoría capaz de unificar los comportamientos; pero a su vez, esa hipótesis se basa en dos hipótesis: la hipótesis de la unificación de los comportamientos y la hipótesis del fisicalismo de los comportamientos. Esas hipótesis se enuncian de la siguiente manera.

*Hipótesis de la unificación de los comportamientos:* Todos los comportamientos de sistemas biológicos y no biológicos forman parte de un único conjunto de fenómenos.

*Hipótesis del fisicalismo de los comportamientos:* Todos los comportamientos que exis-

ten en la realidad pueden ser explicado en términos físicos.

Asumir las dos hipótesis anteriores implica que todos los comportamientos deben poder ser explicados únicamente en términos físicos. Ese tipo de explicación es la que presenta la teoría general de los exocomportamientos.

Otra meta de la tesis doctoral ha sido desarrollar métodos computacionales que contribuyan a la creación robots multifuncionales capaces de integrarse en la vida cotidiana de los seres humanos. Así, también se han desarrollado investigaciones aplicadas que han requerido hipótesis tecnológicas. Concretamente las hipótesis han sido: la hipótesis del cognitivismo basado en condiciones de verdad y la hipótesis de los robots multifuncionales mediante nociones topológicas. Estas dos hipótesis hacen las siguientes afirmaciones:

*Hipótesis del cognitivismo basado en condiciones de verdad:* Las condiciones de verdad son elementos capaces de permitir el desarrollo de métodos que doten de todo tipo de comportamientos cognitivos a robots y computadoras.

*Hipótesis de los robots multifuncionales mediante nociones topológicas:* Se pueden desarrollar métodos cualitativos para la toma de decisiones y comunicación de información que permitan la creación de robots multifuncionales mediante el uso de nociones topológicas

La combinación de esas dos hipótesis ha dirigido la investigación aplicada a buscar representaciones cualitativas del espacio mediante nociones topológicas.

## 0.6. La teoría general del exocomportamiento

El problema de encontrar una teoría fisicalista capaz de unificar los comportamientos se ha abordado investigando y desarrollando la teoría general del exocomportamiento. La teoría general del exocomportamiento se ha desarrollado buscando evitar los problemas que se han observado en teorías anteriores para explicar los comportamientos y que posea las características positivas que han tenido propuestas previas. Así, es interesante destacar que se evitan los conceptos mentalistas y se formulan desde un punto fisicalista. Una consecuencia inesperada de la formulación es que ha llevado a determinar que procesos ontogénicos y los comportamientos pertenecen al mismo conjunto de fenómenos, los exocomportamientos. Una característica destacable de la teoría general del exocomportamiento es que su formulación plantea una completa integración de los exocomportamientos con los procesos evolutivos.

Además del papel de la teoría general del exocomportamiento como explicación, esta

ha servido también de marco teórico en el que plantear preguntas sobre la naturaleza. Así, se han planteado preguntas sobre la evolución que han llevado a realizar una investigación sobre los modelos computacionales de los organismos multicelulares y las fuerzas evolutivas que los han conformado.

Así, la teoría general del exocomportamiento ha mostrado las dos características que se espera de una teoría de unificación: explicar fenómenos y dirigir nuevas investigaciones.

## 0.7. El programa robots multifuncionales mediante nociones topológicas

El problema de desarrollar métodos computacionales cualitativos que contribuyan a la creación de robots multifuncionales ha sido abordado proponiendo y desarrollando el programa de robots multifuncionales mediante nociones topológicas. Este programa de investigación tiene como objetivo desarrollar métodos cualitativos mediante el uso de nociones topológicas con las que definir relaciones espaciales cualitativas. Además del uso de nociones topológicas, se han usado elementos de la teoría cognitiva de condiciones de verdad. El programa afronta el desarrollo de métodos computacionales para desenvolverse en entornos y situaciones de manera gradual. Primero se intenta encontrar un método capaz de afrontar la situación más sencilla posible. Una vez que se ha obtenido un método efectivo, se plantea una situación más compleja, y se estudia si el método es efectivo o no. Si no lo es, se estudian cuáles son los problemas y se busca un nuevo método que sea efectivo y que cumpla con las condiciones que establece el programa. Así, este proceso se repite continuamente. Se debe notar que el programa de robots multifuncionales mediante nociones topológicas está en concordancia con la teoría general del exocomportamiento, ya que las soluciones que se han encontrado para contribuir en la construcción de robots computacionales encajan con las propiedades que se consideran para los exocomportamientos de los seres humanos, y el objetivo del programa de robots multifuncionales mediante nociones topológicas se plantea como la búsqueda de una solución particular a la ecuación general de la teoría general del exocomportamiento.

## 0.8. Resultados y conclusiones

En base a las hipótesis de la investigación, el objetivo de la tesis ha sido investigar y desarrollar la teoría general del exocomportamiento. Ese objetivo se basa en dos motivos. Un motivo es el tener un marco para lograr una mejor comprensión de los comportamientos y sus efectos en la naturaleza. El segundo motivo es tener una guía en el desarrollo de nuevos métodos que doten a computadoras y robots de comportamientos

complejos que les permitan integrarse en entornos humanos no específicos facilitando la vida cotidiana o ayudando en tareas peligrosas. Las investigaciones fundamentales de esta tesis han arrojado varios resultados. La revisión bibliográfica ha mostrado que la inteligencia y la cognición son conceptos que no pueden lograr la unificación, ya que hay comportamientos en la naturaleza que no son el producto de procesar información. En la investigación y desarrollo de la teoría general del exocomportamiento se ha alcanzado una formulación más sólida. Uno de los descubrimientos más significativos de estudiar su formulación ha sido el encontrar que los comportamientos y los procesos ontogénicos forman parte del mismo conjunto de fenómenos naturales, ya que tienen la misma causa. Además, la formulación ha llevado a investigar otras cuestiones de las que se han obtenido resultados importantes. Uno de ellos es que la robustez computacional es una ventaja evolutiva para los sistemas nerviosos de los organismos multicelulares. Otro de los resultados importantes, es que las semánticas de teoría de modelos pueden ser considerados procesos computacionales y que sin ellas no se puede describir completamente el proceso computacional que llevan ciertos sistemas. Además, se propuso un nuevo mecanismo para interpretar dinámicas cerebrales que permitiría la representación de categorías.

En el caso de las investigaciones aplicadas, los resultados han mostrado que la línea de investigación de la creación de robots multifuncionales mediante nociones topológicas y condiciones de verdad es muy prometedora. Se ha logrado una arquitectura, denominada arquitectura de navegación cualitativa topológica 1.0, que permite la navegación en espacios desconocidos y dinámicos en los que las tomas de decisiones, los objetivos, y el conocimiento que almacena se hace de manera cualitativa. Un resultado interesante que se ha obtenido es un método para codificar información direccional en relaciones topológicas. Además, ese resultado ha servido para crear una heurística que permite tomar decisiones sobre giros y definir un sistema de coordenadas cualitativas esféricas que elimina algunas limitaciones que tenía el sistema de coordenadas cualitativas cartesiano construido al comienzo de la investigación. El último resultado que se ha obtenido es una técnica para generar expresiones de lenguaje natural sobre relaciones espaciales cualitativas que abre la puerta a una nueva fase para continuar investigando el desarrollo de robots multifuncionales basados en nociones topológicas y condiciones de verdad. Los resultados obtenidos en esta tesis llevan a concluir que tanto la teoría general del exocomportamiento como el programa de investigación de robots multifuncionales mediante nociones topológicas son sólidas líneas de investigación en la obtención de resultados para la unificación de los comportamientos y la creación de robots multifuncionales.

# Agradecimientos

La tesis doctoral que aquí se presenta es el resultado de no solo mis años de doctorando sino de dos décadas buscando respuestas sobre cómo generar comportamientos y qué son los comportamientos. La lista de gente que me ha acompañado en mi camino por encontrar respuestas es muy numerosa y agradezco a todos ellos el tiempo que me han dedicado. Pero entre todos quiero dar las gracias a los que en este caso me han ayudado o apoyado directamente en mi realización de esta tesis doctoral.

En primer lugar, quiero dar las gracias a mis padres, María Dolores Tomé Martín y Julián Miguel Escolar, por su apoyo incondicional y su financiación para mi investigación. Sin ellos esta tesis nunca hubiera sido posible.

A mis directores de tesis, Ángel Luis Sánchez Lázaro y Luis Alonso Romero, por toda su ayuda para que haya podido realizar y presentar esta tesis doctoral; pero sobre todo por creer en mi investigación y darme la oportunidad de realizarla aceptando dirigirla.

A Rodolfo Llinás por ser siempre un ejemplo para mí de buscar respuestas en la naturaleza, y por acogerme en su departamento para realizar mi estancia en Nueva York permitiéndome discutir la teoría general del exocomportamiento con él y aprender un poco de sus enormes conocimientos sobre el sistema nervioso.

A Moritz Müller por aceptar que realizara mi estancia en el Kurt Gödel Research Center con él y por su tiempo enseñándome complejidad computacional.

A Ángel F. Porteros Herrero por sus consejos y ayuda en mis investigaciones referentes al sistema nervioso.

A Lori-Ann Tuscan que siempre me ayuda corrigiendo mis textos en inglés y en mis cuestiones referentes a psicología

A Evelia Lizárraga Olivar por su ayuda revisando erratas en esta tesis.

A María Gracia Manzano por compartir conmigo sus enormes conocimientos en lógica matemática.

A Luis de la Ossa Jiménez por su interés en mis investigaciones sobre la teoría general del exocompotamiento, ya que siempre ha sido un elemento de ánimo para continuar con ellas.

A la fundación Caja de Burgos por las becas económicas que tuvo a bien concederme para ayudarme a realizar mi estancia en Nueva York y para asistir a congresos.

A Jaime Migue Tomé, mi querido hermano, y que me ha ayudado encontrado erratas perdidas.

# Capítulo 1

## Introducción

El objetivo principal de la ciencia es lograr dar una explicación a los fenómenos que ocurren en el universo material. La explicación científica de un conjunto de fenómenos naturales concreto se realiza a partir de a un conjunto de proposiciones que se ajustan a los datos empíricos que se conocen sobre esos fenómenos y ponen en relación ciertos observables que comparten todos los fenómenos del conjunto. En función de lo amplio que es el conjunto de fenómenos y el número de proposiciones que abarca una explicación científica diferenciamos entre distintos tipos de conjuntos de proposiciones. Así, se habla de ley, modelo o teoría científica. Cada explicación científica no solo es capaz de explicar fenómenos, sino también de producir predicciones de resultados experimentales.

La ciencia aumenta el conocimiento que se posee sobre la naturaleza creando una teoría para un conjunto de fenómenos para el que no poseía una explicación o reduciendo el número de conjuntos de fenómenos independientes. El proceso por el cual el número de conjuntos de fenómenos considerados independientes queda reducido es denominado unificación. El resultado de un proceso de unificación es una teoría de unificación, la cual explica un conjunto de fenómenos que está constituido por la unión de conjuntos de fenómenos que eran considerados hasta ese momento independientes. A veces la ciencia ha buscado lograr una teoría que unifique determinados conjuntos de fenómenos y en otras ocasiones la teoría ha surgido inesperadamente. El problema fundamental para encontrar una teoría de unificación es formular la equivalencia entre las causas de los conjuntos de fenómenos que se quieren unificar. La mecánica de Newton, la teoría celular, la teoría cinética de los gases, las leyes de Maxwell o la relatividad general son teorías científicas que han unificado diferentes conjuntos de fenómenos de la naturaleza.

Un conjunto de fenómenos que busca explicar la ciencia es el de los comportamientos. El intento de explicar los comportamientos se remonta al menos hasta la Antigua Grecia. Hasta el siglo XX un comportamiento es el fenómeno por el que las entidades biológicas interactúan con su entorno. Sin embargo, desde la aparición de los compu-

tadores y robots, los comportamientos son también fenómenos que son producidos por sistemas no biológicos. El primer intento por unificar los comportamientos de animales y máquinas fue propuesto por Norbert Wiener en base al concepto de retroalimentación; aunque solo obtuvo un éxito parcial, su propuesta cambió la manera de ver los comportamientos. Actualmente, distintas disciplinas científicas como la psicología, la etología, la neurociencia, la inteligencia artificial, la robótica o la lingüística estudian distintos conjuntos de comportamientos. A pesar de ser un tema que ha recibido mucha atención por parte de la comunidad científica, los comportamientos siguen siendo una de las cuestiones fundamentales de la ciencia básica porque no se tiene una teoría general para abordar los comportamientos. Además, los descubrimientos en este campo tienen implicaciones en la creación de tecnologías que ayuden a crear robots o sistemas que puedan integrarse en la vida diaria de los seres humanos. Así, encontrar una teoría que unifique los comportamientos es uno de los grandes retos de la ciencia.

## 1.1. Motivación

Esta tesis doctoral tiene como principales objetivos el investigar y desarrollar una teoría de unificación de los comportamientos, así como nuevos métodos destinados a crear robots multifuncionales. Así, las motivaciones de esta tesis doctoral pertenecen tanto a la investigación fundamental como a la investigación aplicada. En toda investigación fundamental, la principal motivación que hay detrás de ellas es la curiosidad. El estado actual de la ciencia obliga a considerar los comportamientos de los microorganismos, los animales, los vegetales, los seres humanos y los de robots o computadoras por separado. Así, uno se pregunta: ¿son realmente fenómenos diferentes?, ¿qué causa los comportamientos? o ¿podría un robot desarrollar comportamientos como los de los seres humanos? Además, las investigaciones en el campo de la evolución han mostrado que los comportamientos están involucrados en fenómenos evolutivos. Lo que plantea preguntas tales como ¿existen distintos tipos de comportamientos desde el punto de vista de los procesos evolutivos? ¿pueden influir esos tipos en los procesos evolutivos? ¿se pueden hacer predicciones sobre fenómenos evolutivos a partir de las causas de los comportamientos? Para dar respuestas que sacien nuestra curiosidad a todas esas preguntas y otras muchas más, es necesario poseer una teoría de unificación de los comportamientos. Así, la investigación sobre una teoría científica que explique el conjunto entero de los comportamientos desde la investigación básica tiene un doble motivo. El primero es lograr la unificación de los comportamientos, ya que una teoría de unificación supone un enorme paso hacia una mejor comprensión de la naturaleza. El segundo es integrar los comportamientos dentro de la teoría de la evolución de manera que se tenga una teoría general de como los comportamientos se relacionan con los fenómenos evolutivos.

Pero toda nueva teoría científica abre una nueva puerta a la investigación aplicada.

En este caso, una teoría general de los comportamientos tiene especial interés en los campos de la inteligencia artificial y la robótica. Así, otra de las motivaciones de esta tesis doctoral es tener una teoría que sirva de marco para desarrollar algoritmos y estructuras de datos que permitan a robots y computadoras interactuar de manera natural con el ser humano y llevar a cabo comportamientos útiles para el ser humano. La aparición de tecnologías que doten a robots y computadores de la capacidad de realizar comportamientos más complejos tendría un impacto directo en la sociedad y la industria.

## 1.2. Hipótesis de trabajo

Debido a que entre las investigaciones que se han llevado a cabo, unas pertenecen a investigación fundamental y otras a investigación aplicada, las hipótesis que se han empleado también se dividen en hipótesis fundamentales e hipótesis tecnológicas. A continuación se prestara cada uno de los dos conjuntos de hipótesis y sus implicaciones para la investigación.

### 1.2.1. Hipótesis fundamentales

La hipótesis principal detrás de la investigación que se ha llevado a cabo es la hipótesis de que la teoría general del exocomportamiento es una teoría capaz de unificar los fenómenos de los comportamientos. Pero esa hipótesis envuelve otras dos hipótesis, la hipótesis de la unificación de los comportamientos y la hipótesis del fisicalismo de los comportamientos:

*Hipótesis de la unificación de los comportamientos:* Todos los comportamientos de sistemas biológicos y no biológicos forman parte de un único conjunto de fenómenos.”

Hay un hecho que hace sospechar que esta hipótesis es verdad. Los comportamientos, independientemente de que provengan de organismos cuyos comportamientos actualmente se consideran que pertenecen a conjuntos diferentes, provocan la aparición de los mismos fenómenos evolutivos. Por ejemplo, castores y seres humanos son capaces de modificar su entorno, lo que tiene como consecuencia que se eliminen presiones evolutivas y se inhiba la evolución. La hipótesis de la unificación de los comportamientos contiene en realidad dos unificaciones. Una es la unificación entre los comportamientos de los seres humanos y los del resto de seres vivos, ya que siempre se ha considerado que los comportamientos de los seres humanos eran completamente diferentes de los del resto de seres vivos. La otra unificación que contiene es entre los comportamientos de los seres vivos y de los comportamientos de sistemas no biológicos autónomos. Hay que notar que esta hipótesis no implica que el conjunto final en el que se unifiquen los

comportamientos este formado únicamente por esos fenómenos. Por ejemplo, cuando Maxwell llevo a cabo la unificación de los fenómenos eléctricos y magnéticos se vio que los fenómenos ópticos también formaban parte del mismo conjunto.

La segunda hipótesis que se asume es la hipótesis del fisicalismo de los comportamientos. La hipótesis afirma lo siguiente:

*Hipótesis del fisicalismo de los comportamientos:* Todos los comportamientos que existen en la realidad pueden ser explicado en términos físicos.

Esta segunda hipótesis niega que cualquier clase de dualismo intervenga en la generación de comportamientos en la naturaleza. Es decir, todos los comportamientos deben poder ser explicados únicamente en términos físicos.

### 1.2.2. Hipótesis tecnológicas

La investigación aplicada que se ha llevado a cabo para desarrollar métodos para la robótica ha tenido dos hipótesis en las que se han fundamentado. Una es la hipótesis que se asume en la teoría cognitiva de las condiciones de verdad.

*Hipótesis del cognitivismo basado en condiciones de verdad:* Las condiciones de verdad son elementos capaces de permitir el desarrollo de métodos que doten de todo tipo de comportamientos cognitivos a robots y computadoras.

La segunda hipótesis tecnológica que se ha asumido es la que ha dado a lugar al programa de robots multifuncionales mediante nociones topológicas.

*Hipótesis de los robots multifuncionales mediante nociones topológicas:* Se pueden desarrollar métodos cualitativos para la toma de decisiones y comunicación de información que permitan la creación de robots multifuncionales mediante el uso de nociones topológicas

Dadas estas dos hipótesis, la investigación se ha dirigido a estudiar la definición de condiciones de verdad para nociones topológicas y cómo estas son aplicables en el desarrollo de métodos que puedan crear robots multifuncionales que se integren en la vida diaria.

### 1.3. Objetivos de la investigación

En base a las hipótesis de la investigación, el objetivo de la tesis ha sido continuar investigando los fundamentos de la teoría general del exocomportamiento para usarla como marco tanto para abordar una mejor comprensión de los comportamientos y sus efectos en la naturaleza, como para guiar el desarrollo de nuevos métodos que doten a computadoras y robots de comportamientos complejos que les permitan integrarse en entornos humanos no específicos facilitando la vida cotidiana o ayudando en tareas peligrosas. Ese objetivo se puede descomponer concretamente en la siguiente serie de objetivos:

- Estudiar la formulación y el alcance de la teoría general del exocomportamiento. La propuesta de la teoría general del exocomportamiento se encontraba en un estado muy incipiente al comienzo. Así, su formulación debe ser analizada y discutida buscando evitar ambigüedades o limitaciones. También es necesario abordar en el análisis su poder explicativo y predictivo, mediante la consideración de las leyes que posee para explicar la relación entre la causa que propone y fenómenos de la naturaleza.
- Estudiar la interpretación física microscópica de la causa propuesta por la teoría general del exocomportamiento para los comportamientos. Dado que la teoría general del exocomportamiento se basa en la hipótesis del fisicalismo de los comportamientos es necesario, investigar como se interpreta la causa propuesta por la teoría general del exocomportamiento por la física.
- Analizar mediante la TGE cuestiones paradójicas en las ciencias cognitivas. Una de las razones por las que se adopta una nueva teoría científica es porque resuelve paradojas que aparecen entre los hechos que muestra la naturaleza y las teorías actuales. Por eso, dentro de los objetivos, uno de ellos es estudiar las aportaciones de la TGE a resolver paradojas.
- Abordar cómo los comportamientos y su generación influyen en los procesos evolutivos. La teoría de la evolución es una de las teorías científica más sólida que se tiene en la ciencia. Sin embargo, existe una diversidad de fenómenos para los que no se ha generado una explicación, a veces porque algunas cuestiones no se han planteado. Una teoría que unifique los comportamientos tiene que enmarcarse dentro de la teoría de la evolución dando explicación y prediciendo procesos evolutivos.
- Desarrollar nuevos métodos para la robótica enmarcados en la TGE. Una teoría científica además de permitir explicar fenómenos, es valiosa por la capacidad de llevar a cabo investigación aplicada para generar o mejorar tecnologías como consecuencia de la mejor comprensión de los fenómenos que explica.

## 1.4. Evolución de la investigación

El trabajo presentado en esta tesis doctoral se inició en el año 2010 tras la realización de mi trabajo fin de master en el Master de Sistemas Inteligentes de la Universidad de Salamanca en el que se refinaba la formulación inicial de la teoría general del exocomportamiento. Las investigaciones que se han llevado a cabo durante el periodo del doctorado se han desarrollado a cabo de manera bastante paralela. Especialmente ha habido mucha independencia entre las investigaciones fundamentales y las aplicadas, ya que los resultados de unas no afectaban a las otras. Así, se comenzaron a la vez las investigaciones sobre la formulación de la teoría general del exocomportamiento y las aplicadas para el programa de robots multifuncionales mediante nociones topológicas al mismo tiempo. La evolución de las investigaciones fundamentales ha sido la siguiente. Primero se abordó la formulación de la teoría general del exocomportamiento. En el análisis de la formulación se descubrieron ciertas situaciones paradójicas que hicieron reconsiderar la formulación de la teoría. Una versión corta de la formulación de la teoría general del exocomportamiento fue presentada en la International Work-conference on the Interplay between Natural and Artificial Computation 2011 [370]. Una vez se tuvo una formulación más sólida, se abordó el análisis de la teoría desde el punto de vista físico, buscando conocer si la teoría no entraba en contradicción con la física y si la propia física podía ser una base sólida para la teoría. Abordadas esas cuestiones se exploró cómo la teoría general del exocomportamiento puede abordar la descripción de un observador y los argumentos del campo de filosofía de la mente en contra del paradigma computacional de las ciencias cognitivas. Tras esa investigación, se empezó a considerar la relación entre la teoría general del exocomportamiento y el sistema nervioso de los animales. A partir de ahí, se han desarrollado dos investigaciones distintas. Una, sobre cómo el sistema nervioso es capaz de codificar estados simbólicos que ha llevado a proponer un nuevo mecanismo para explicar las dinámicas de actividad cerebral. Este resultado fue presentado en la International Work-conference on the Interplay between Natural and Artificial Computation 2015 [376]. La segunda, sobre la explicación evolutiva de por qué en los animales solo se encuentra el modelo computacional de red neuronal en sus sistemas nerviosos. Los resultados de este último se han publicado en la revista *Natural Computing* [377].

La evolución de las investigaciones aplicadas comenzó por un profundo análisis del estado del arte de las técnicas de navegación en robótica, con especial atención en los métodos cualitativos. Además, se hizo un profundo análisis de las relaciones topológicas relativas y la heurística de semántica topológica cualitativa. Este trabajo se publicó en los artículos en el congreso International Work-conference on the Interplay between Natural and Artificial Computation 2013[373] y en un artículo en la revista *Connection Science* [379]. Después se investigó la extensión de la heurística de semántica topológica cualitativa para que contemplara situaciones diferentes. Así, se estudiaron

dos opciones. Una fue hacer una extensión incorporando relaciones de cercamiento. La otra extensión que se estudio, fue el usar modificadores borrosos. Los resultados de esa investigación se publicaron también en un artículo en la revista *Robotics and Autonomous Systems* [374]. Tras esto, se abordó el diseño de una arquitectura que permita la navegación autónoma en espacios desconocidos y dinámicos abiertos que hiciera uso de la heurística de semántica topológica cualitativa. Los resultados de la arquitectura han sido publicados en un artículo en la revista *Connection Science* [378]. Una vez diseñada y probada la arquitectura en un entorno virtual, se abordó una de las limitaciones que poseía el empleo de la heurística de semántica topológica cualitativa, la toma de decisiones sobre giros. La investigación encontró un método de codificar información direccional en información topológica definiendo un conjunto de relaciones denominadas relaciones topológicas direccionales. En base a esas relaciones se ha creado, el método denominado heurística de semántica cualitativa direccional que permite la toma de decisiones sobre giros. Además, usando las relaciones topológicas direccionales y las 8 relaciones topológicas clásicas, se ha propuesto combinarlas para crear un sistema de coordenadas cualitativas alternativo al que permite definir las relaciones topológicas relativas. Una vez se han desarrollado métodos suficientes como para permitir que un robot tenga la autonomía suficiente para navegar, se ha establecido un plan de investigación para usar las nociones topológicas y la teoría de cognitiva de condiciones de verdad y se ha comenzado la investigación de métodos para la generación de lenguaje natural.

## 1.5. Esquema general

La investigación que se presenta en esta tesis doctoral ha estado enfocada en estudiar la teoría general del exocomportamiento como una teoría de unificación de los comportamientos de sistemas biológicos y no biológicos autónomos. Por lo tanto, la teoría general del exocomportamiento ha sido el objetivo de las propias investigaciones o el marco en el que se han concebido y que las ha guiado. Así, algunos capítulos tienen ciertas dependencias de otros. Aun así, diversos resultados de los que se han obtenido tienen interés por ellos mismos porque son relevantes para cuestiones que se abordan en disciplinas concretas.

La organización de esta tesis doctoral es la siguiente. El capítulo 2 hace una revisión del problema de explicar los comportamientos desde el nacimiento de la ciencia y cuáles son las características y cuáles los problemas que han surgido con las distintas teorías que se han propuesto. El capítulo 3 hace una pequeña revisión de la teoría de la evolución y de la importancia que tienen los comportamientos para esta teoría y los procesos que explica. Tras los dos anteriores capítulos, el capítulo 9 presenta la teoría general del exocomportamiento como propuesta para unificar los comportamientos y la nueva formulación que se propone de ella.

En el capítulo 5 se aborda cuál es la interpretación física microscópica de la explicación que se propone en la teoría general del exocomportamiento. Así, el capítulo contiene una revisión de la noción de computación e información en el campo de la física, y una discusión de la importancia de estos conceptos en la física fundamental.

El capítulo 6 aborda cuál es el método de representación de estados simbólicos en el sistema nervioso y cómo ha surgido desde el punto de vista evolutivo. En ese análisis para abordar las cuestiones se explican los métodos que se han propuesto hasta ahora para interpretar las dinámicas cerebrales y además se propone un nuevo mecanismo, denominado Trayectorias-Lenguaje-Estado, que podría ser usado por el cerebro para codificar estados simbólicos.

El capítulo 7 aborda principalmente por qué el modelo computacional que se encuentra en el sistema nervioso de los animales es el mismo en todos. En el capítulo se hace un profundo análisis de los procesos y factores que pueden haber influido en la evolución en el sistema nervioso de organismos multicelulares. El capítulo muestra que la robustez computacional es una ventaja evolutiva, y que el modelo de red neuronal ha sido seleccionado porque la posee.

El capítulo 8 aborda la formulación del paradigma computacional, el proceso de observación y las limitaciones del paradigma computacional aseveradas por lógicos y filósofos de la mente en el marco de la teoría general del exocomportamiento. El análisis que se lleva a cabo muestra que las semánticas creadas por Tarski deben ser consideradas procesos computacionales, y por lo tanto, que deben de tenerse en cuenta para llevar a cabo una descripción computacional completa de un sistema. Esta conclusión es fundamental para los siguientes capítulos.

El capítulo 9 hace una presentación de los conceptos de la teoría cognitiva de condiciones de verdad. La teoría cognitiva de condiciones de verdad es un marco matemático que permite describir agentes y las condiciones de verdad que hay en sus procesos de representación del entorno y tomas de decisiones. En base a las conclusiones del capítulo anterior, la teoría cognitiva de condiciones de verdad se propone para describir los tipos de valores de la causa que según la teoría general del exocomportamiento permitirían la creación de robots multifuncionales.

El capítulo 10 presenta el programa de investigación denominado robots multifuncionales mediante nociones topológica y toda la investigación que se ha realizado en este programa desarrollando métodos nuevos que puedan ser usados en la creación de robots multifuncionales. Esas investigaciones se han guiado mediante la formulación de la teoría cognitiva de condiciones de verdad presentada en el capítulo anterior. La mayoría de las secciones están dedicadas a métodos que se han desarrollado para abordar la navegación cualitativa. La última sección del capítulo presenta la planificación que se propone para investigar la generación y procesamiento de lenguaje natural y los primeros resultados en la generación de lenguaje natural que se han obtenido. Finalmente, el capítulo 11 presenta las conclusiones de la investigación que se ha llevado a cabo y expone las futuras líneas de investigación con las que continuar el trabajo que

se ha llevado hasta ahora.

Las relaciones entre los capítulos se muestran en la figura 1.1.

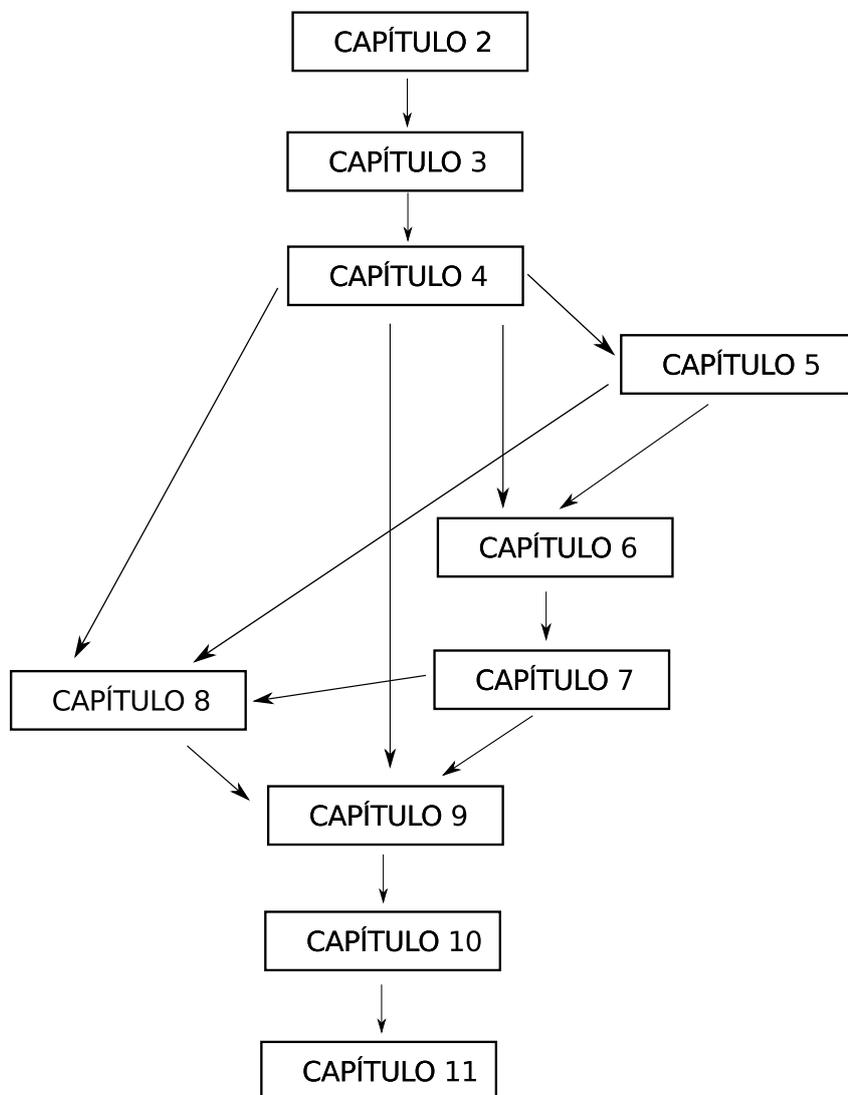


Figura 1.1: Relaciones de dependencia entre los capítulos que componen la tesis doctoral



## Capítulo 2

# Revisión histórica de las propuestas para explicar los comportamientos

Este capítulo revisa los orígenes, los descubrimientos, la evolución de las teorías para explicar los comportamientos y su estado actual. Hay dos hechos que se observan durante la revisión. La primera es que todo apunta a que los distintos comportamientos pueden ser considerados fenómenos de un mismo conjunto. La segunda que la inteligencia es un concepto problemático que se autocontradice en las distintas definiciones que se han propuesto de él y que debería ser abandonado para encontrar una teoría que unifique los comportamientos.

### 2.1. Explicaciones de los comportamientos hasta el siglo XIX

Las teorías para explicar los comportamientos se remontan al menos hasta la Antigua Grecia [623]. La concepción de que los comportamientos son de dos tipos, instintivos e inteligentes, se basa fundamentalmente en las ideas elaboradas por Platón para explicar los comportamientos de los seres vivos. Platón proponía que los comportamientos de los hombres de los que carecen los animales eran generados por el alma racional y el resto por el alma irracional. Al carecer los animales y plantas del alma racional, estos solo tendrían comportamientos vegetativos y orgánicos. Así, los comportamientos que se denominan inteligentes serían los que Platón proponía como los generados por el alma racional y los instintivos los generados por el alma irracional. Veinte siglos después de Platón, Descartes mantenía las ideas de Platón, ya que también consideraba que los seres humanos y los brutos (animales) eran diferentes debido a que el ser humano tenía alma [123]. Así, según él, los brutos eran máquinas con complejos mecanismos para controlar su comportamiento, pero carecían de alma racional que sí tenía el ser humano. Esa alma racional, según él, era lo que permitía que el hombre tuviera lenguaje

y símbolos para comunicarse y lo que hacía que la inteligencia le permitiera adaptarse al entorno. En cambio, el comportamiento de los brutos sería generado innatamente sin intervención de la inteligencia. Esta visión se mantuvo hasta la llegada de Charles Darwin con el mecanismo de la selección natural para explicar los comportamientos instintivos e inteligentes sin el uso de ningún tipo de entidades místicas o religiosas [115]. En un principio, Darwin intentó explicar los comportamientos instintivos mediante el hábito, pero se encontró con que no era posible explicar muchos comportamientos mediante ese proceso. Sin embargo, vio que el mecanismo de la selección natural sí los explicaba. Así, Darwin propuso que los comportamientos instintivos son el resultado de la variación de otros comportamientos instintivos mediante selección natural. Respecto a los comportamientos inteligentes, Darwin también propuso que su aparición ha sido causada por la selección natural. Recientemente Steven Pinker retomó la cuestión de cómo la selección natural ha dirigido la evolución hasta crear a los seres humanos con las capacidades que conocemos de generar complejos comportamientos [463]. La teoría de la evolución determina que el comportamiento es un rasgo más del fenotipo que está sujeto a la selección natural, y el comportamiento de los animales es un antecedente del que posee el hombre entre los que existe una continuidad. Así, descarta la necesidad de un “diseñador” que determine los comportamientos de cada organismo y determinar que comportamientos son extinguidos y cuales seleccionados, pero no explica los mecanismos que generan los comportamientos <sup>1</sup>. Para intentar explicar los comportamientos, estos fueron divididos en tres grandes grupos, los instintivos, los hábitos y los inteligentes, y se consideró que los instintivos son causados por el instinto, los hábitos por la repetición y los inteligentes por la inteligencia.

Con la llegada de la teoría de la evolución el interés por encontrar explicaciones científicas para los comportamientos creció rápidamente, aunque el comienzo por desarrollar explicaciones científicas de los mecanismos de los comportamientos pueden establecerse en la propuestas de Albrecht von Haller[203] frente a la teoría neuroeléctrica de Tommaso Laghi [283]. Partiendo de ahí, repasemos brevemente algunos descubrimientos que se realizaron para explicar cómo se generan los comportamientos entre los siglos XVII y XIX. En un principio la teoría de Haller se impuso y prácticamente se desestimó la teoría de Lagui. Sin embargo, la teoría neuroeléctrica fue retomada por Luigi Galvani. Las investigaciones experimentales de Galvani entre 1780 y 1799 le llevaron a formular la hipótesis de que la electricidad estaba envuelta en la conducción nerviosa y la contracción muscular, convirtiéndose en el fundador de la electrofisiología [76]. También es interesante mencionar que en 1729 el astrónomo francés Jean-Jacques d’Ortous de Mairan informó de que el movimiento de las hojas de algunas plantas emergía de ritmos endógenos [332]. En el siglo XIX, hay que destacar la propuesta de Franz Joseph Gall de que las funciones mentales de los humanos se encontraban en

---

<sup>1</sup>Aunque la teoría de la evolución si puede explicar por qué la naturaleza usa unos mecanismos y no otros al determinar una ventaja evolutiva en el mecanismo seleccionado

regiones específicas del cerebro. Esa hipótesis recibió un fuerte respaldo en 1861 con la investigación de Paul Broca en la que localizaba el centro del habla en el cerebro. Hay que destacar también los trabajos de Ivan Mikhaylovich Sechenov sobre reflejos y cerebro que fueron centrales para que Ivan Pavlov llevara a cabo sus investigaciones experimentales que probaron la teoría de reflejos condicionados [486]. En 1869, Francis Galton continuando con las ideas de Darwin defiende que las capacidades intelectuales de los seres humanos son innatas y están sujetas a la selección natural al igual que los rasgos físicos [169]. Al final del siglo XIX, Ramón y Cajal revolucionó la investigación del sistema nervioso aportando las primeras pruebas de la existencia de células nerviosas, su teoría neuronal y la ley de la despolarización, que le convirtieron en el padre de la neurociencia moderna [91].

## 2.2. Investigaciones de los comportamientos durante la primera mitad del siglo XX

A pesar del cambio de considerar entidades místicas a entidades y procesos físicos como causas de comportamientos, a principios del siglo XX la causa de los comportamientos instintivos y los inteligentes se consideraban completamente diferentes y era lo que diferenciaba a los humanos del resto de los seres vivos [605]. En el año 1906 Charles Scott Sherrington publicó su teoría que la capacidad integradora del sistema nervioso central era debida a un conjunto de reflejos coordinando actividades de distintos órganos [103]. Respecto a la distinción entre instinto e inteligencia, esta empezó a ser criticada [411]; pero los trabajos de Alfred Binet sobre la medida de resolución de problemas por humanos y las posteriores investigaciones de Charles Spearman y Thurstone Louis dieron una aparente solidez científica a la propuesta de la existencia de la inteligencia. Así, la psicometría empezó a ser tomada seriamente. En el campo del comportamiento animal, Edward Thorndike, basándose en su investigación, enunció la *ley del efecto* en 1911. Las investigaciones de Thorndike fueron la base para la posterior formulación y desarrollo del condicionamiento operante por Burrhus F. Skinner. En 1926 Walter Cannon retomó el concepto de Claude Bernard de estabilidad del medio interior, y propuso una propiedad de los sistemas que regula su medio interno, la homeostasis [92]. En 1927 Karl von Frisch publicó sus investigaciones sobre el sistema de comunicación animal que tienen las abejas obreras para transmitir la dirección y distancia de la fuente de polen. En los años 30 comenzaron sus investigaciones Konrad Lorenz y Tinbergen Nikolaas sobre el instinto, proponiendo los patrones de acción fija [319]. También en los años 30, Nicolas Rashevsky investigó y propuso el primer modelo matemático de red neuronal [478]. En 1938, Skinner publicó su teoría sobre el comportamiento que inició las investigaciones sobre conductismo [533]. En 1943, Warren McCulloch y Walter Pitts propusieron la modelización matemática de una red neuronal

en un contexto booleano [360]. Su artículo comenzó la teoría de las máquinas de estados finitos y las redes neuronales artificiales. Su propuesta es importante porque identifica el sistema nervioso con un sistema computacional. También en 1943, Kenneth Craik propuso la unificación de cerebros y máquinas [110]. Desde ese momento, el lograr esa unificación ha sido un importante objetivo para la ciencia. Por su parte, Jean Piaget estaba investigando las etapas de la inteligencia y su desarrollo haciendo descubrimientos muy importantes [460]. Las investigaciones de Piaget, a diferencia de las de psicometría, identificaban la inteligencia como una característica común de todos los individuos humanos. En 1948, Edward Chace Tolman propuso los mapas cognitivos. La propuesta de Tolman permite explicar los complejos comportamientos observados por Wolfgang Köhler en chimpancés que no se pueden explicar por la ley de efecto [270], y preparó el camino para la revolución cognitiva que comenzaría en los años 50. Un año después, Norbert Wiener definió cibernética como el estudio del control y la comunicación en los animales y las máquinas, proponiendo que animales y máquinas operan de acuerdo a los mismos principios [620]. En 1949, Donald Hebb formuló su teoría de la asamblea celular, ahora conocida como teoría hebbiana, combinando psicología y fisiología [215].

Dados todos los descubrimientos citados, emergió la visión de que los comportamientos estaban divididos en distintos conjuntos para los que se tenían teorías. Específicamente se consideraron siete tipos de comportamientos donde cada uno se explicaba independientemente del resto. Esos ocho tipos fueron: los comportamientos homeostáticos, ritmos biológicos [539], reflejos, patrones de acción fija, comportamientos por habituación/sensibilización, comportamientos por condicionamiento clásico, comportamientos operantes, y comportamientos inteligentes.

- **Comportamientos homeostáticos.** Estos comportamientos sucederían fundamentalmente en todos los organismos, y serían efectuados para mantener una condición interna estable a pesar de los cambios que sucedan en su entorno. A pesar de que el objetivo es mantener una condición interna estable, el comportamiento consiste en un intercambio regulado de materia y energía con el exterior. Los comportamientos de las células y microorganismos serían todos de este tipo.
- **Ritmos biológicos.** Estos fenómenos han sido estudiados durante siglos, pero los mecanismos que los producen han empezado a ser explicados solo con el reciente desarrollo de la biología molecular .
- **Reflejos.** Son los comportamientos simples que aparecen siempre ante el mismo estímulo. El reflejo ha sido el primer mecanismo neuronal de generación de comportamientos en ser investigado.
- **Patrones de acción fija.** Estos comportamientos son más complejos que los reflejos, pero permiten muy poca variación en las causas que los desencadenan.

- **Comportamientos por habituación/sensibilización** Son comportamientos modificados por la manera más sencilla que hay de apredizaje.
- **Comportamientos por condicionamiento clásico.** Estos comportamientos son activados por un estímulo, pero a diferencia de los reflejos, el estímulo que desencadena el comportamiento puede ser cambiado pudiendo el organismo determinar un nuevo estímulo como desencadenante.
- **Comportamientos operantes.** Son los comportamientos que son controlados por sus consecuencias.
- **Comportamientos inteligentes.** Son los comportamientos complejos llevados a cabo por el ser humano que no están en ninguno de los conjuntos anteriores. Inicialmente se pensó que podrían ser explicables mediante condicionamiento operante pero los experimentos que mostraron la existencia de representaciones mentales descartaron la posibilidad de que fueran causados por condicionamiento operante. A partir de ese momento surgió la psicología cognitiva que ha estado intentando explicarlos mediante propuestas de procesamiento de la información.

De los ocho conjuntos de comportamientos el único para el que no se lograba una teoría que los explicara es el de los inteligentes. Tanto los intentos del conductismo [550], la psicología del desarrollo [459, 460] y la neuropsicología [215] no lo consiguieron. Pero la escala de Binet y Simon y las posteriores investigaciones en psicometría soportaban la teoría de la inteligencia. La teoría de la inteligencia basada en los trabajos de Binet y Simon puede condensarse en tres postulados principales:

1. Hay un conjunto cerrado de comportamientos autónomos llamados comportamientos inteligentes.
2. Los comportamientos inteligentes son causados por una característica o propiedad de los sistemas, la inteligencia.
3. Las diferencias individuales en los comportamientos inteligentes vienen causadas por las diferencias en la propiedad de la inteligencia.

Dada esa situación, el mayor reto que existía en los años 50 en el estudio de los comportamientos era explicar la inteligencia.

### 2.3. Explicaciones de la inteligencia a partir de la segunda mitad del siglo XX

La explicación de los comportamientos en la segunda mitad del siglo XX empieza un nuevo período con la revolución cognitiva [382]. La demostración científica de que

los estados mentales existen mediante la medición de rotaciones mentales derrumbó los estrictos dogmas del conductismo para investigar y explicar los comportamientos [430]. Hoy sabemos que el conductismo no es capaz de explicar los denominados comportamientos inteligentes y por lo tanto tampoco sería capaz de explicar qué es la inteligencia [581]. La psicología cognitiva, a diferencia del conductismo, asume que los procesos mentales implicados en el conocimiento son métodos de procesamiento, almacenamiento y recuperación de información y por lo tanto investigables científicamente.

La revolución cognitiva estuvo íntimamente relacionado con la aparición de las computadoras. La posibilidad de tener una máquina que es capaz de calcular cualquier función efectivamente calculable hizo surgir una nueva disciplina, la Inteligencia Artificial (IA) [354]. La IA mostró desde sus inicios que las computadoras pueden generar comportamientos complejos [383]. El objetivo de la IA desde su inicio ha sido hacer que los computadores generen comportamientos inteligentes [532]. Obviamente, la IA podría alcanzar su objetivo si la inteligencia fuera una función computacional específica. Así, bajo la hipótesis de que la inteligencia es una función computacional, la IA ha buscado encontrar cuál es esa función para generar comportamientos inteligentes. Además, el hecho de que la psicología cognitiva considere que los procesos cognitivos consisten en procesamiento de la información hizo que la IA fuera su vertiente experimental. La existencia de computadoras y de la IA ha provocado que las teorías cognitivas no puedan ser solo explicaciones verbales, se deben implementar computacionalmente para verificarse. Esto ha obligado a una clarificación conceptual de los mecanismos que proponen las teorías cognitivas. Como resultado de estas conexiones el objetivo tanto de la psicología cognitiva como de la IA durante la segunda mitad del siglo XX ha sido explicar la inteligencia.

Durante más de un siglo casi cada científico que ha investigado los comportamientos ha tenido su propia definición de inteligencia. Aun así, todas las definiciones tienen ciertas características generales que comparten. Una de ellas es que o bien se ha considerado a la inteligencia una propiedad extensional o una intensional. El enfoque intensional se enfoca en encontrar qué es la inteligencia porque considera que solo hay una causa capaz de generar los comportamientos inteligentes. El extensional se enfoca en generar comportamientos inteligentes porque considera que hay muchas causas que los pueden generar. Así, desde el punto de vista extensional basta con encontrar una causa capaz de generarlos y no hace falta encontrar la que los causa en los humanos. La discusión entre extensional e intensional ha surgido tanto en lo referente a los constituyentes físicos que componen el sistema como a la función computacional que lleva el sistema.

Otro elemento que ha sido muy importante en las definiciones ha sido el paradigma computacional, aunque en este caso la cuestión del paradigma computacional está más relacionada con la metodología para encontrar la función computacional que sería la inteligencia que con la definición de la inteligencia.

Dado que la inteligencia que se presupone a los seres humanos no ha podido ser explicada, se ha empezado a hablar de inteligencia computacional, que causaría com-

portamientos inteligentes, pero sería una propiedad diferente de la inteligencia humana y que sería poseída por las máquinas.

A continuación, se presenta una clasificación de los más destacados intentos para explicar qué es la inteligencia o para generar comportamientos inteligentes que se han llevado a cabo hasta el momento. En ella no se contemplan las propuestas de que la inteligencia no es explicable o reproducible artificialmente por no ser una propiedad física. En la clasificación de las propuestas se distinguen cinco clases particulares: inteligencia simbólica intensional, inteligencia subsimbólica intensional, inteligencia simbólica extensional, inteligencia subsimbólica extensional e inteligencia física intensional. Cada categoría simbólica y subsimbólica puede dividirse en blanda y dura. Dentro de las duras no se considera que la inteligencia requiera manejar inexactitud o incertidumbre. En el caso de las blandas, se ha considerado que una parte fundamental del comportamiento humano es la capacidad de manejar inexactitud e incertidumbre. Así un método se clasifica como blando si los valores o representaciones permiten manejar valores con incertidumbre o inexactitud, y duro de lo contrario. La clasificación se muestra en la figura 2.1. Esta clasificación es tan buena como cualquier otra que se pueda proponer, pero esta nos es útil para posteriores discusiones en esta tesis por las tres variables que se manejan en ella.

Las dos categorías más generales son inteligencia físicamente intensional y físicamente extensional. La categoría de propuestas que determinan la inteligencia como una propiedad físicamente extensional, se conoce en filosofía de la mente como funcionalismo. El funcionalismo determina que cualquier sistema con capacidad de computación igual a una máquina de Turing podría ser capaz de tener inteligencia. La inteligencia físicamente intensional se divide a su vez en diferentes categorías. Por otra parte, también se ha propuesto que la inteligencia sea físicamente intensional, pero esta propuesta no surgió hasta que se vio que los avances en IA estaban por debajo de la progresión que se esperaba. Así, se propuso que habría propiedades físicas específicas que no todos los sistemas físicos poseen, que permitirían calcular funciones que no pueden ser calculadas por una máquina de Turing.

A continuación, se mencionarán propuestas que han sido realizadas para explicar qué es la inteligencia en cada una de las categorías particulares de la clasificación propuesta.

### **2.3.0.1. La inteligencia como propiedad intensional simbólica**

Esta categoría fue el primer planteamiento que se utilizó para intentar hallar qué es la inteligencia y cómo generar comportamientos inteligentes. Esta categoría se basa en dos condiciones. Una es considerar que sólo existe una inteligencia y otra que el marco adecuado para describirla es el paradigma simbólico. Los principios del paradigma simbólico fueron establecidos por Allen Newell y Herbert Simon en su disertación al obtener el premio Turing en 1975 [419]. En su disertación enunciaron la hipótesis del

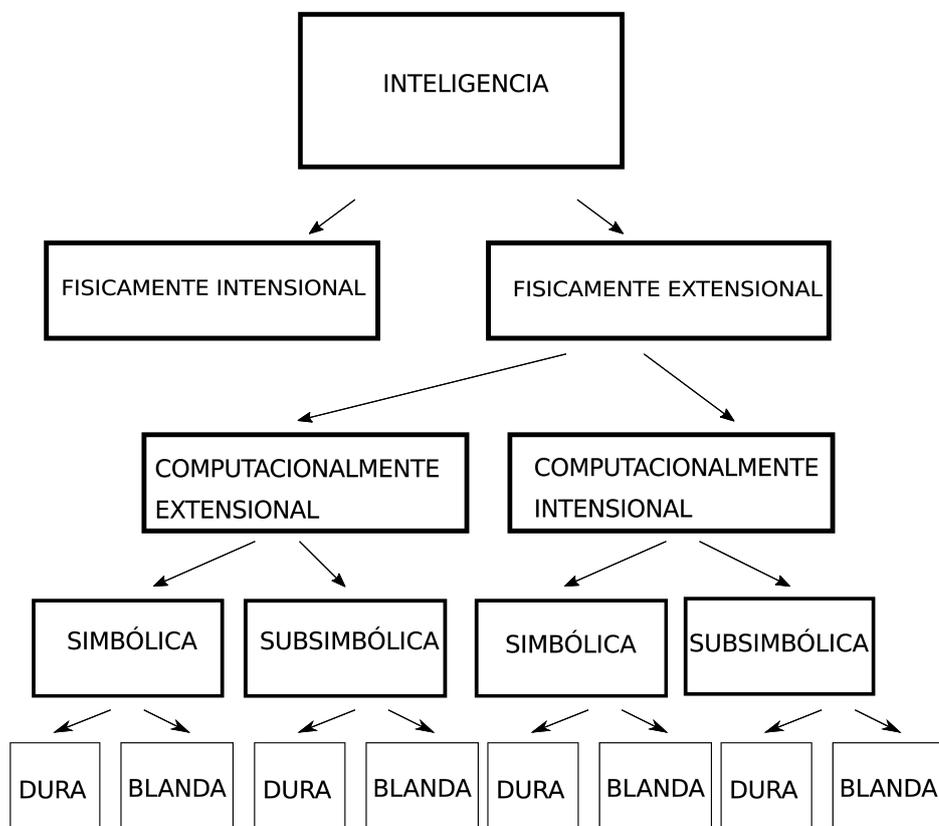


Figura 2.1: El esquema de la figura muestra una clasificación de propuestas para la explicar en qué consiste la inteligencia

sistema de símbolos físicos y la hipótesis de la búsqueda heurística. La hipótesis del sistema de símbolos físicos la enunciaron de la siguiente manera:

“The Physical Symbol System Hypothesis. A physical symbol system has the necessary and sufficient means for general intelligent action.” [416] p.116

Es decir, la hipótesis determina que el dominio de objetos en el que la IA debe centrarse para investigar el fenómeno de la inteligencia. Ellos lo circunscriben a los sistemas de símbolo físico, a los que definen de la siguiente manera.

“A physical symbol system consists of a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called an expression (or symbol structure).” [416] p.116

Esa definición hace que el sistema tenga que obedecer las leyes de la física, por lo

que se trata de un planteamiento fisicalista monista.

La hipótesis de la búsqueda heurística la formularon de la siguiente manera:

“Heuristic Search Hypothesis. The solutions to problems are represented as symbol structures. A physical symbol system exercises its intelligence in problem solving by search—that is, by generating and progressively modifying symbol structures until it produces a solution structure.” [416] p.120

En esta hipótesis, Newell y Simón propusieron el tipo de procesos en los que se daría la inteligencia. Para ellos, la inteligencia es un proceso que evita tener que hacer una búsqueda exhaustiva en todo el espacio de búsqueda para encontrar una solución; pero además dejan claro que existe una estructura de símbolos y que el sistema realiza operaciones que transforman unos símbolos por otros hasta que la estructura es una solución al problema que se le plantea.

El paradigma simbólico nació antes que Newell y Simon formularan sus hipótesis, ya que sus hipótesis nacieron del trabajo experimental que ellos y otros investigadores habían llevado a cabo. En los primeros años de la IA emergieron cuatro líneas de investigación para intentar explicar que es la inteligencia. Una es la de Newell y Simon, otra la de John McCarthy y las dos últimas fueron propuestas por Marvin Minsky. Newell y Simon comenzaron a considerar que los modelos computacionales podían ser teorías de los procesos cognitivos [416]. La segunda es la de McCarthy que proponía el uso de modelos basados en sistemas lógicos para entender la inteligencia [355]. Esa línea de investigación se ve hoy en los trabajos sobre conjeturas de Enric Trillas y sus colaboradores [580, 579]. Por su parte, Minsky consideraba que la propuesta de McCarthy iba a fallar, y que además, era necesario manejar contenidos más grandes y estructurados de lo que lo hacía la lógica para explicar la inteligencia, y propuso la teoría de los marcos [386]. Sin embargo, algunos años después Minsky hizo una nueva propuesta para explicar la inteligencia. Su propuesta fue que la mente humana es una sociedad enorme de agentes que procesan el conocimiento [384]. Su propuesta fue el origen de lo que ahora se denominan sistemas multiagentes [309].

Respecto al intento de crear teorías computacionales de procesos cognitivos, en 1973 Newell se dio cuenta que las teorías computacionales que se proponían eran para características muy específicas y que se daba la situación de que esas teorías llegaban a estar en contradicción entre ellas, lo que él llamó el problema de las microteorías [417]. Así, Newell propuso la necesidad de que se debían proponer teorías que unificaran las teorías que se habían creado para explicar la cognición de los seres humanos. La propuesta de Newell fue muy sensata porque como ya se ha mencionado uno de los métodos por los cuales la ciencia avanza es la creación de teorías de unificación. En 1983, John R. Anderson defendió la idea de que todos los procesos cognitivos de alto nivel deben ser sustentados por el mismo sistema y propuso que ese sería la *ACT\** [22]. Dado que habría un único mecanismo, él defendió que existiría una única teoría de la mente

para explicar la cognición (que contendría a la inteligencia). En 1990, Newell enfatizó la idea de crear un programa de desarrollo de teorías de la unificación que llevara a encontrar la teoría que fuera capaz de explicar todos los fenómenos cognitivos [414]. En su programa de investigación, cada teoría de unificación debería ser instanciada en una arquitectura cognitiva <sup>2</sup>. Comparando el funcionamiento de las arquitecturas cognitivas en computadoras con lo que hace el ser humano mostraría cuál de ellas es correcta. Este proceso debería conducir a encontrar, la teoría unificada de la cognición. El programa propuesto por Newell ha sido bien aceptado por la comunidad científica. Desde su propuesta muchos investigadores han participado en desarrollar arquitecturas cognitivas y lo continúan haciendo [276, 499, 286]. Hay arquitecturas cognitivas como SOAR en la que han trabajado más de 100 personas.

Paradójicamente, el éxito de las arquitecturas es también su fracaso. En contra de lo que Newell pensaba, unas teorías no han ido descartándose frente a otras en la explicación de procesos cognitivos por lo que el número de arquitecturas cognitivas ha ido creciendo.

### 2.3.0.2. La inteligencia como propiedad intensional subsimbólica

La IA comenzó usando el enfoque simbólico para intentar generar comportamientos inteligentes, pero aunque los comienzos fueron prometedores pronto encontró grandes dificultades en lograr realizar tareas que en teoría no deberían ser complejas por la gran cantidad de animales que las llevan a cabo pese a no tener altas capacidades cognitivas. Una alternativa al enfoque simbólico fue el procesamiento subsimbólico mediante redes neuronales artificiales, y que se ha denominado el enfoque conexionista. Aunque McCulloch y Pitts habían dado en 1943 el primer modelo computacional de las redes neuronales, el enfoque conexionista no empezó a tomar protagonismo hasta el año 1986. Ese año David E. Rumelhart, Geoffrey E. Hinton y R. J. Williams dieron la regla generalizada delta [495]. La regla delta generalizada fue el primer método para lograr aprendizaje en redes neuronales con capas ocultas. Esto fue clave porque las redes con capas ocultas no tienen las limitaciones computacionales que sí tienen las redes neuronales sin capas ocultas. Se había demostrado que las redes sin capas ocultas solo pueden separar clases que sean linealmente separables [385]. Tras esto, Rumelhart y James L. McClelland propusieron abordar la comprensión de la cognición a través de modelos computacionales de redes neuronales. Los resultados logrados por Rumelhart y sus colaboradores tuvieron tanta importancia que a partir de los años 80 la ciencias cognitiva abandonaron el marco del procesamiento de información como fundamento para tomar el conexionismo. Así, comenzó una nueva etapa para explicar y generar comportamientos inteligentes [349, 389].

---

<sup>2</sup>Actualmente, el termino arquitectura cognitiva es usado como sinónimo de teoría unificada de la cognición

Actualmente existe todo un campo dedicado a la investigación de la computación en el cerebro, denominado neurociencia computacional. El aumento en las capacidades computacionales de las computadoras comerciales está permitiendo que se investigue cómo el cerebro computa [90, 21] y están apareciendo nuevas propuestas [638]. Los descubrimientos hechos sobre cómo el cerebro humano son usados para lograr comportamientos humanos en computadoras [438, 352]. Al mismo tiempo se han usado los modelos conexionistas para intentar explicar la inteligencia de los seres humanos [568]. Todas las propuestas de esta categoría buscan explicar los comportamientos mediante la comprensión del funcionamiento de los sistemas nerviosos de animales y el ser humano. Sin embargo, a pesar de los numerosos resultados obtenidos, nos encontramos muy lejos de explicar el comportamiento cognitivo en base a procesos computacionales llevados a cabo por las neuronas [256, 177]. Este hecho proviene de que el número de neuronas y conexiones que tiene el cerebro humano junto a las propiedades que posee cada neurona [312] hacen que crear computaciones de igual complejidad esté lejos de nuestros medios disponibles de computación.

### 2.3.0.3. La inteligencia como propiedad extensional simbólica

El enfoque extensional tiene sus raíces en la definición de Turing de inteligencia [587], ya que él fue el primero en considerar que si un sistema genera comportamiento inteligente indistinguible del de un ser humano entonces se debe admitir que tiene inteligencia. La diferencia de esta categoría con la intensional simbólica es que no se considera que existe una única inteligencia, ya que ambas comparten el paradigma simbólico como marco.

Este nuevo planteamiento tiene que ver con el surgimiento de la ingeniería del conocimiento en la que se busca simplemente tener éxitos puntuales generando un comportamiento en algún ámbito concreto y no importa si el cómo se genera es igual o distinto a cómo los seres humanos lo generan. La aparición de la ingeniería del conocimiento se debió a que la búsqueda de una función o algoritmo capaz de dotar de inteligencia humana a una computadora no se había conseguido. Así, la ingeniería del conocimiento intenta construir sistemas que resuelven problemas concretos, y se usa el mejor método computacional para obtener un comportamiento concreto sin que se considere que un método u otro es la causa de la inteligencia [71, 512, 494]. Algunos de los métodos que se usan son: lógica clásica, lógica modal, lógica descriptiva, lógica no-monotónica, lógica borrosa, probabilidad o redes bayesianas.

Más recientemente, han surgido otras propuestas para generar comportamiento inteligente como AIXI[237] y la máquina de Gödel [508]. AIXI es una teoría de decisión secuencial desde una perspectiva de la teoría algorítmica de la información. La máquina de Gödel es un resolvidor de problemas autoreferenciales que hace auto mejorar su código cuando puede probar que el cambio mejora el código anterior. Estas propuestas no claman que explican la inteligencia sino que son capaces de generar comportamientos

inteligentes.

#### **2.3.0.4. La inteligencia como propiedad extensional subsimbólica**

En esta categoría la inteligencia es determinada por la dinámica de las interacciones entre los elementos del sistema. No se considera que exista una sola dinámica que sea en si misma la inteligencia sino que existen multiples. En este planteamiento acoge distintas propuestas desde las arquitectura de subsunción [85] a redes que procesan mediante principios diferentes de los que enuncia la neurociencia. En el caso de redes artificiales, la máquina de boltzman [7] o el algoritmo de retropropagación del error [111] forman parte de esta categoría porque no importa que no tengan bases biológicas sino que son capaces de generar comportamientos como los que se buscan.

#### **2.3.0.5. Inteligencia intensional física**

La situación continuada de la IA de no alcanzar su objetivo principal junto con las expectativas iniciales de que sería una meta que se alcanzaría rápidamente hizo que se planteara si el problema es que las computadoras comerciales carecen de propiedades físicas que sí tiene el cerebro humano y que permitirían procesos de computación que no son posibles en las computadoras comerciales. Concretamente esas propiedades estarían relacionadas con procesos cuánticos. Uno de los proponentes ha sido Roger Penrose [449]. Penrose ha propuesto que las altas capacidades cognitivas de los seres humanos son consecuencia de las leyes cuánticas. Así, la razón por la que la IA fracasa en lograr su meta residiría en las limitaciones del hardware de nuestras computadoras actuales. Solo una nueva generación de ordenadores cuánticos podría lograr inteligencia general. Sin embargo, se ha demostrado que los argumentos matemáticos de Penrose son erróneos [78]. Pero, aun así, Penrose y Stuart Hameroff han desarrollado una teoría sobre cómo podría ser llevada a cabo computación cuantica por el sistema nervioso de los seres humanos [204]. Si la teoría de Hameroff and Penrose es posible está discusión. Se ha argumentado que las condiciones físicas necesitadas son razonablemente improbables que puedan darse en el cuerpo humano [563].

## **2.4. Descubrimientos sobre los comportamientos de distintos organismos**

En la anterior sección se ha mencionado que la revolución cognitiva comenzó una nueva etapa en la explicación de los comportamientos y la inteligencia mediante descripciones computacionales, pero la revolución cognitiva también trajo consigo mucho mayor interés por investigar el comportamiento de los animales. También, desde el punto de la teoría de la evolución se empezó a hacer más incapie en que la inteligencia no

puede ser una característica desligada del resto de las características que pueda tener el resto de organismos. David Stenhouse propuso en una teoría sobre como la inteligencia habría evolucionado a partir de mecanismos de instinto [545, 546]. Por otro lado, los avances acaecidos en electrofisiología y biología molecular a partir de la segunda mitad del siglo XX han permitido adentrarse mucho más en los mecanismos físicos de los comportamientos de animales, plantas y microorganismos. Los descubrimientos que se han sucedido han hecho que la visión sobre los comportamientos que puede realizar cada organismo y la división de los comportamientos que se tenía entorno a la segunda mitad del siglo XX ha ya cambiado completamente. A continuación, se revisan descubrimientos que se han realizado desde la segunda mitad del siglo XX y han demostrado que dogmas establecidos anteriormente sobre los comportamientos y los organismos estaban equivocados.

### 2.4.1. Cognición animal

La revolución cognitiva de la psicología también tuvo su reflejo en el estudio de los comportamientos de los animales. Así, nació el campo de la cognición animal que estudia el proceso mediante el cual un animal recibe información de su entorno y la procesa. Dentro de la cognición animal se encuentra la cognición comparativa que estudia y compara procesos básicos como la percepción, categorización, memoria y aprendizaje entre las especies. Algunas de las preguntas más importantes que la investigación en cognición animal ha abordado desde su aparición han sido las siguientes [54]:

- ¿Son los animales capaces de tener algo semejante de algún modo a las capacidades lingüísticas que tienen los humanos?
- ¿Qué entienden los animales de sus mentes y las mentes de otros?
- ¿Tienen los animales capacidades metacognitivas?
- ¿Tienen los animales capacidades de categorización?
- ¿Tienen los animales capacidades numéricas?
- ¿Tienen los animales memoria a largo y corto plazo?
- ¿Tienen los animales memoria episódica?
- ¿Tienen los animales memoria espacial?
- ¿Tienen los animales habilidades sociales?

Los resultados de las investigaciones realizadas sobre las anteriores preguntas no han indicado que las capacidades de los seres humanos sean radicalmente diferentes de las que existen en animales, sino más bien que las capacidades de los animales son un antecedente de las que tienen los seres humanos [133, 596]. A continuación, se mencionan algunos de los descubrimientos más relevantes desde la revolución cognitiva que han afectado directamente a nuestra manera de diferenciar los comportamientos de animales y seres humanos.

Una de las características que se ha considerado única del hombre es el lenguaje, considerándose que la capacidad de comunicación de los animales era inferior a la de los seres humanos. Uno de los primeros intentos en poner a prueba esa tesis fue realizado por William Furness en 1916. Furness intentó enseñar durante algunas horas a un orangután hembra a hablar inglés, pero sólo pudo pronunciar pocas palabras. Posteriormente, Keith y Catherine Hayes criaron a un chimpancé como si fuera un niño humano pero el chimpancé aprendió también solo unas pocas palabras [213]. El problema principal de esos intentos por enseñar a simios a hablar parece ser que tiene que ver con cuestiones anatómicas [131]. Para evitar esos problemas, Allen y Beatrice Gardner realizaron investigaciones sobre la capacidad de aprender el lenguaje de signos por chimpancés [173]. Este nuevo enfoque mostró un panorama completamente diferente. El chimpancé era capaz de usar 132 signos que incluían nombres, pronombres y verbos. Esos signos eran capaces de combinarlos en series de hasta 5 signos. Tras esos resultados, el lenguaje de signos fue usado por más investigadores, todos con resultados positivos. El más sistemático entre ellos fue Herber Terrace [566]. En esa investigación el chimpancé fue capaz de aprender 125 signos que los uso para crear más de 19.000 expresiones.

Otras investigaciones han mostrado que no solo los hominoideos son capaces de comprender y emplear símbolos, también monos capuchinos son capaces de usar fichas como símbolos [10]. En el caso de los chimpancés se ha comprobado que ellos tienen capacidades de metacognición [53]. Incluso se ha encontrado que los perros tienen capacidades cognitivas como memoria episódica [163].

Pero las primeras investigaciones que empujaron a repensar los dogmas sobre las capacidades cognitivas de los animales fueron realizadas por Irene Peperberg con el loro gris Alex [451, 450, 455]. La demostración de que un loro gris es capaz de clasificar por categorías y reconocer cantidades de hasta seis era realmente tan inesperado en ese momento que atrajo la atención de los medios de comunicación [453]. Las investigaciones con loros grises han continuado hasta nuestros días mostrando capacidades que se creían reservadas solo a humanos o primates evolutivamente muy cercanos a los seres humanos [452, 456]. Pero los loros grises no son las únicas aves con capacidades que no se esperaban. Las investigaciones con cuervos han arrojado también resultados inesperados [218, 5]. Incluso investigaciones con el gallo bankiva han mostrado que estos realizan comunicación multimodal [535] .

Hay que mencionar que los descubrimientos de la existencia de mecanismos de

aprendizaje en numerosos vertebrados hizo que Euan Macphail planteara la hipótesis nula [328, 329]. La hipótesis nula propone que todos los vertebrados tienen los mismos mecanismos de inteligencia y que el hombre solo se diferencia del resto en su capacidad para el lenguaje. Por inverosímil que pueda parecer inicialmente la hipótesis, lo cierto es que no se ha logrado refutar [448]

Otras investigaciones que han afectado a dogmas establecidos sobre los comportamientos de los animales son los llevados a cabo con invertebrados. Se ha considerado que la ratio de neocortex está correlacionada con el grado de inteligencia [296]. Pero, la organización del cerebro de los invertebrados difiere profundamente de la de los vertebrados. Entre los invertebrados las investigaciones han mostrado que los cefalópodos tienen altas capacidades cognitivas [190], siendo los pulpos su máximo exponente [487, 278]. Esto ha demostrado que el tamaño, o la ratio, del neocortex no puede considerarse como criterio único para determinar las capacidades cognitivas.

También se debe resaltar la investigación sobre medusas, ya que las medusas no exhiben cefalización, solo poseen una red nerviosa. Pese a que estos animales carecen de cerebro, algunas medusas tienen una gran capacidad para detectar la dirección de las corrientes del océano y nadar fuertemente en contra de ellas. Eso ha hecho que se cuestione la idea tradicional de qué es un cerebro [504, 175, 327, 231]. Estudiando las cubomedusas se hizo un descubrimiento sorprendente. Las cubomedusas son nadadoras ágiles y fuertes y predadores activos [503]. Las cubomedusas se diferencian de otras medusas en que tienen un aparato sensorial muy elaborado y demuestran un comportamiento activo como el de los peces. La cubomedusa posee un sistema visual sorprendente, que está formado por una matriz interactiva de 24 ojos de cuatro tipos distintos. Los experimentos han mostrado que las lentes de las cubomedusas contienen un gradiente de índice de refracción ajustado finamente que produce imágenes que están muy cerca de estar libres de aberraciones [427]. Aunque, los experimentos también han mostrado que la posición de la retina no coincide con la imagen nítida, lo que conlleva campos receptivos muy amplios y complejos en fotoreceptores individuales [427, 611]. Esos descubrimientos implican que la arquitectura de los ojos es útil en servir a una única tarea visual. Lo que convierte a la cubomedusa en un ejemplo de que la visión no es un sistema de propósito general.

### 2.4.2. Las plantas y sus comportamientos

Los descubrimientos en el reino de las plantas han trastocado también nuestra visión de las capacidades que se adjudicaban a los distintos organismos en función de su distancia al hombre. Las investigaciones realizadas han llevado a proponer que las plantas son organismos inteligentes [576, 575, 577]. Las investigaciones han mostrado distintos tipos de comportamientos y complejos mecanismos.

Un tipo de comportamiento de las plantas que se ha investigado en profundidad son los ritmos circadianos. Estos han sido investigados fuertemente durante los últimos

30 años [208, 357]. Los ritmos circadianos regulan aspectos de crecimiento y desarrollo, y aumentan la adecuación biológica de las plantas. Los seres humanos hemos tenido conocimiento de la existencia de ritmos circadianos desde el siglo cuarto a.C., pero no fue hasta 1729 que la ciencia comenzó a dar cuenta de ellos [332]. Aún así, no se ha empezado a considerar a las plantas como organismos genuinamente capaces de desarrollar comportamientos hasta finales de los años 80 del siglo pasado [531]. Los avances en biología molecular han permitido adquirir una profunda comprensión de los ritmos circadianos. Nuestra actual comprensión de los ritmos circadianos establece que son endógenos y causados por relojes circadianos, que en plantas están mayormente basados en factores de transcripción que se inhiben mutuamente entre ellos. Aun así, el mecanismo ha mostrado ser mucho más complejo de lo que se pensó inicialmente. Los relojes circadianos son un mecanismo de predicción que permite al organismo anticipar cambios regulares en el entorno. Los relojes circadianos controlan muchos procesos en las plantas. Uno de los comportamientos que controlan son los movimientos de las hojas. Ese control por medio de relojes circadianos da una mayor adecuación biológica a la planta en determinados entornos. Así, el movimiento de las hojas por ritmos circadianos ha sido un rasgo seleccionado. En investigaciones realizadas se ha mostrado que plantas del género *Arabidopsis* son capaces de aprender los ciclos de luz y oscuridad, lo cual aporta también mayor adecuación biológica [192]. La capacidad de aprendizaje implica predicción porque la planta que aprende puede adaptar sus predicciones a diferentes duraciones de luz y oscuridad causadas por las estaciones del año. Se ha descubierto en *Arabidopsis* que las plantas también son capaces de calcular el ritmo al cual las reservas de almidón fotosintético acumuladas durante el día pueden ser consumidas para evitar la inanición durante la noche [514, 515].

También se ha encontrado que son capaces de liberar sustancias moleculares volátiles con objetivos como, comunicarse con otras plantas, atraer animales polinizadores, repeler herbívoros, o atraer enemigos de herbívoros [232]. Incluso se ha descubierto que son capaces de detectar las vibraciones que producen los herbívoros al alimentarse de la planta para producir defensas químicas, y diferenciarlas de las causadas por el viento o el canto de insectos [23].

Entre las plantas vasculares sensitivas destaca la mimosa y la venus atrapamoscas. La venus atrapamoscas, que es una de las plantas más estudiadas, tiene hojas que le permiten capturar insectos, y ha desarrollado la realización de dos predicciones para ser efectiva [631]. Primeramente, debe decidir si lo que está entre las hojas es comida, o no, para prevenir cerrarse con arena o cualquier otro elemento que no es útil. Para distinguir objetos inanimados de presas usa como marcador recibir dos estímulos mecánicos consecutivos con una diferencia de tiempo entre ellos menor de 20 a 30 segundos. Cuando la venus atrapamoscas detecta los dos estímulos, salta a un estado de semicierre. Tras llegar al estado de semicierre, debe decidir si la presa merece la pena para cerrarse completamente. El mecanismo para tomar esta decisión funciona de la siguiente manera: si el insecto es pequeño, entonces podrá escapar a través de los huecos

que deja el estado de semicierre, y la planta no registrará más señales. Sin embargo, si el insecto es del tamaño apropiado no podrá escapar del estado de semicierre y producirá estímulos mecánicos constantemente al intentar escapar. Al registrar esos estímulos, la trampa se cerrará completamente. Así, la venus atrapamoscas es capaz de predecir el tamaño de su presa. La venus atrapamoscas salta del estado de abierta al estado de semicerrada en 0,03 segundos, y del estado de semicierre a completamente cerrada en 0.03 segundos [631]. Esta velocidad es lograda mediante el uso de señales eléctricas. Las células de la venus atrapamoscas y otras plantas sensitivas pueden generar potenciales de acción que permiten los tiempos de reacción mencionados.

Ese hecho nos lleva a mencionar que se ha descubierto que las plantas no solo usan señales eléctricas para disparar movimientos rápidos de hojas en plantas sensitivas, sino que son usadas por procesos fisiológicos en plantas ordinarias [630, 162]. Las plantas envían señales eléctricas a distancias cortas y largas. Como en los animales, las plantas tienen tres tipos de señales eléctricas, potenciales de acción, potenciales electrotónicos y potenciales graduados [595]. El mecanismo celular equivalente al sistema nervioso de las plantas nunca desarrolla el mismo grado de complejidad que el sistema nervioso de los animales, pero sí poseen una sencilla red neuronal. Esa sencilla red neuronal está particularmente presente dentro del floema que se ocupa de la comunicación a largas distancias [635, 595, 162]. Las plantas no han desarrollado sinapsis químicas, pero poseen canales alineados equivalentes a uniones en hendidura de los animales, denominados plasmodesmos [490]. Los plasmodesmos permiten la conexión eléctrica entre células y la propagación de los potenciales de acción [552, 162]. Se han hallado evidencias de la transmisión de información mediante potenciales de acción en el árbol de aguacate [440] y también se ha descubierto que las células de las plantas establecen modos de intercambio de información entre ellas que tienen propiedades en común con la sinapsis química de los animales [35, 34].

Otro reciente descubrimiento es que el ácido gamma-aminobutírico (GABA) es, además de un neurotransmisor en animales, una molécula señalizadora en las plantas [476]. Adicionalmente, se ha demostrado que los transportadores de malato activados por aluminio son traductores clave de las señales mediante GABA en las plantas. GABA aparece en las plantas en respuesta a temperaturas extremas, deshidratación, salinidad, stress oxidativo, daño mecánico, acidosis, infecciones víricas y defensa contra herbívoros [66, 267]. No se conoce si la existencia de ligaduras de GABA en múltiples reinos es debido a una evolución convergente o a un mecanismo de GABA ancestral, pero queda claro que las plantas usan GABA para activar determinados comportamientos en respuesta al estado interno o del entorno.

La neurobiología de las plantas ha emergido como una disciplina debido a los descubrimientos de señales electroquímica en respuesta a estímulos del entorno [75, 440]. Entre las líneas de investigación, la más sorprendente es una dedicada a estudiar la hipótesis de “raíz-cerebro” propuesta por primera vez por Charles y Francis Darwin [36]. Los Darwin propusieron que las raíces se comportan como lo hacen los animales infe-

riores con su ápice asentado en el polo anterior del cuerpo de la planta donde actuaría como si se tratará de un cerebro. Partiendo de descubrimientos recientes, la propuesta de los Darwin ha sido reconsiderada y se ha propuesto que el ápice de la raíz es una estructura equivalente en las plantas al cerebro [33, 31]. Específicamente, se ha descubierto que en la zona de transición del ápice de la raíz recibe información sensorial de la cofia e indica las respuestas motoras de las células en la zona de elongación [33]. La hipótesis de la “raíz-cerebro” ha sido avalada por test de decisiones binarias hechas por raíces de maíz en laberintos de tipo Y [632]. Además, han sido registradas dinámicas espacio-temporales de la actividad eléctrica de la red en el ápice de la raíz, lo que ha llevado a proponer que su función es integrar señales internas y externas para desarrollar adaptaciones a los cambios del entorno [343].

Toda la investigación que se ha citado muestra que las plantas son organismos sensoriales y comunicativos, caracterizados por comportamientos activos de resolución de problemas para adaptarse a su entorno.

### 2.4.3. Los hongos y sus comportamientos

Los hongos son organismos que también han mostrado comportamientos inesperados. Las investigaciones realizadas han revelado que tienen relojes circadianos que predicen ciclos de la naturaleza. El hongo brasileño *Neonothopanus gardneri* exhibe una luminiscencia excepcionalmente intensa que atrae a los insectos, que ayudan a dispersar sus esporas. El transporte de las esporas por los insectos provee de un efectivo medio de dispersión y garantiza una ventaja al fungi especialmente en bosques densos. Se ha demostrado que la bioluminiscencia fúngica es demasiado débil para ser percibida por los animales durante el día, pero es fácilmente percibible por la noche. Por lo tanto, un control rítmico de la bioluminiscencia es un mecanismo que elimina un gasto de energía inefectivo durante el día [435]. El comportamiento del *Neonothopanus gardneri* demuestra que es capaz de predecir el estado del entorno.

### 2.4.4. Los microorganismos y sus comportamientos

Durante el siglo XX se asumió que el comportamiento de los microorganismos era completamente explicado por los principios homeostáticos. Frente a esa visión cabe destacar que en 1979 se demostró que un *Paramecium caudatum* puede ser condicionado clásicamente [220]; pero fue el advenimiento de la biología molecular lo que hizo que empezara a considerarse la capacidad de los microorganismos de hacer complejos comportamientos predictivos. Todo se inició cuando el genoma de diversos microorganismos fue secuenciado. Al hacerlo se encontró que tenían una enorme cantidad de componentes genéticos. Eso incrementó el interés en entender cómo los genes y proteínas interactúan en redes celulares complejas, lo que hizo aparecer el campo de la computación molecular [318].

La investigación teórica ha mostrado que podrían evolucionar para predecir relaciones entre estímulos y situaciones ambientales adaptándose al orden temporal de su aparición [239], y podrían realizar aprendizaje asociativo [149]. La investigación experimental ha mostrado que el orden temporal natural de los estímulos está integrado en el cableado de las redes reguladoras de la *Escherishia coli* [396]. La complejidad de procesar estímulos que puede realizar una red biomolecular es tan alta que han sido comparadas con las redes neuronales de los animales [25]. Recientes experimentos han demostrado que *Physarum polycephalum*, un organismo sin neuronas, pueden llevar a cabo habituación, una forma de aprendizaje inconfundible [61]. El número de investigaciones recogiendo comportamientos complejos de los microorganismos ha sido tan importante en los últimos 15 años que se ha propuesto que realmente llevan a cabo procesos cognitivos[324].

### 2.4.5. Comportamientos de apuesta

Uno de los más sorprendentes descubrimientos que se ha hecho es que diversas especies poseen mecanismos de apuesta para tomar decisiones. Las estrategias de apuesta diversificada (Bet-hedging) han sido descritas en insectos y plantas de semillas para afrontar entornos impredecibles [72]. En la escala microscópica, la persistencia bacteriana está relacionada con un mecanismo para decidir el fenotipo estocásticamente [31]. También se han realizado experimentos que han mostrado que los bacteriófago lambda, que están entre los organismos más sencillos, toman decisiones sobre desarrollo, haciendo decisiones estocásticas [29]. Los experimentos realizados sugieren que la conmutación estocástica es una estrategia exitosa de supervivencia en entornos fluctuantes [6]. Estos descubrimientos muestran que la selección natural no siempre elige mecanismos que procesan información del exterior para tomar decisiones porque para algunos entornos apostar es el comportamiento que da mayor adecuación biológica al organismo.

## 2.5. Inteligencia: Un concepto confuso e incapaz de unificar

Durante todo el siglo XX se consideró que explicar qué es la inteligencia era un problema fundamental de la ciencia. Una y otra vez se ha intentado afrontar su explicación redefiniendo el término [548]. Durante la primera mitad del siglo XX, la psicología lo intentó desde diferentes aproximaciones, pero no fue capaz de responder a qué es la inteligencia. La revolución cognitiva y la llegada de las computadoras permitió un nuevo enfoque para intentar responder a qué es la inteligencia. Newell escribió:

“Intelligence is another central concept in cognitive science.” [414]pag. 88

Sin embargo, como él mismo reconoció en la siguiente afirmación, las diferentes definiciones de inteligencia se contradicen entre ellas.

“The notion of intelligence as a scientific construct defined entirely by a technology for creating tests is set in opposition to a notion of defining it by cognitive theories and experiments.” [414] pag.89

Pero, ¿cuál es el motivo de que se apele constantemente a una noción tan problemática? El propio Newell explicaba porqué esa noción no se ha abandonado pese a las contradicciones que posee en su siguiente frase:

“Some notion is required of a graded ability or power of mind like system, which applies in some way to ranges or classes of tasks and ranges or classes of minds to indicate what minds can perform what tasks” [414] pag.89

El concepto de inteligencia se ha usado con tres significados diferentes según Howard Gardner[167]. Uno de los significados referencia a una característica de cada especie. Otro es el de una característica del individuo que determina las diferencias individuales. El tercero es el de los ajustes que lleva a cabo el individuo para ejecutar una tarea. Diferentes líneas de investigación han priorizado alguno de esos tres significados en el campo de la IA. Pero tras más de 60 años de investigación ha sido imposible lograr las metas de explicar qué es la inteligencia y que una máquina exhiba inteligencia general. Así, algunos científicos piensan que la inteligencia es una de las propiedades más misteriosas [71]. En cambio, esta situación ha hecho que otros científicos hayan abierto el debate sobre la necesidad de nuevos conceptos que sustituyan al de inteligencia [395, 634, 82], y se ha afirmado que la inteligencia es un concepto precientífico que debería ser dejado atrás para poder avanzar [395].

La aparición de las técnicas de neuroimagen ha permitido la aparición de un campo en la psicometría que investiga las diferencias individuales de la inteligencia en base a los aspectos anatómicos y bioquímicos [289]. Estas investigaciones han propuesto explicar la inteligencia a través de características neuronales como la hipótesis de la mielinización [325], la hipótesis de la poda neuronal [251] o la hipótesis de la plasticidad [174]. Descubrimientos hechos por técnicas de neuroimagen [251] han creado grandes esperanzas en el campo de la psicometría [202]. Sin embargo, a pesar de que han obtenido ciertos resultados nos encontramos muy lejos de explicar el comportamiento en base al funcionamiento de las neuronas [256]. Pero aun si una teoría general sobre la computación neuronal fuera lograda, no tendríamos una completa comprensión de los comportamientos porque esa teoría está ligada a un modelo de computación específico,

el de red neuronal. En el terreno de psicometría se ha defendido que la inteligencia es lo que mide los test [589, 65]. Pero esa definición implica una cuestión inesperada. Dado que no se ha encontrado relación entre la resolución de problemas complejos y la noción general de inteligencia [615], y teniendo en cuenta la definición de inteligencia la identifica con los test, se deduce que la resolución de problemas complejos no puede ser explicada por la inteligencia. Así, habría que buscar otra causa para explicar la resolución de problemas cognitivos complejos.

Desde las investigaciones de Binet y Simon, tras cada intento por explicar la inteligencia han aparecido problemas con la definición y su aplicabilidad [489, 311, 506]. En este momento, aunque las escalas de la inteligencia tienen utilidad práctica, la psicología considera que la etiqueta “inteligencia” ya no tiene significado científico porque ha sido usada para fenómenos demasiado diversos [188], incluso se ha llegado a proponer que la inteligencia debe ser considerada un mito [506]. Incluso en el mejor de los casos, estamos en la misma situación que hace más de 100 años, no se sabe qué es la inteligencia [188]. Se debate sobre si la inteligencia es mayormente heredable o si puede ser alterada, si hay una inteligencia única o múltiples, si es una esencia pura e inmutable, o una emergente [549].

Junto a los problemas para explicar los comportamientos inteligentes se han realizado muchos descubrimientos que han cambiado nuestros dogmas sobre los comportamientos inteligentes y la inteligencia. Las primeras investigaciones de Wolfgang Köhler y Edgard C. Tolman con chimpancés mostraron que los comportamientos inteligentes podrían no ser solo humanos [270]. Pero el verdadero golpe a los dogmas vino con las investigaciones de Irene Peperberg con loros grises [451] porque la estructura del cerebro de los loros grises difiere de la de los humanos. Más recientemente se han observado comportamientos inteligentes en cuervos [5, 218]. Se han encontrado en plantas comportamientos autónomos que no se esperaban como: termorregulación [523], reacciones al entorno [622], e incluso la capacidad de aprender la duración del día [358]. Otros descubrimientos que han golpeado profundamente los dogmas sobre la inteligencia han sido los descubrimientos de comportamientos de bacterias que no pueden ser explicados en el marco de la homeostasis [239, 396], y que han sido denominados como inteligencia procariota [339]. Estos descubrimientos tienen que ser tenidos muy en cuenta porque las bacterias no tienen sistema nervioso y por lo tanto una teoría neuronal no sería una explicación para estos tipos de comportamientos.

A todo lo anterior, hay que añadir los descubrimientos de comportamientos seleccionados por la naturaleza que nunca se habían considerado como los comportamientos aleatorios [40, 281]. El que los comportamientos aleatorios sean más efectivos que los que procesan información en ciertos entornos derrumba la idea de que los comportamientos inteligentes superan a cualquier otro tipo de comportamiento. Además, se ha descubierto que algunos comportamientos cíclicos también dan mayor adecuación a determinados entornos que comportamientos que procesan información del entorno [358, 193]. Estos últimos descubrimientos son muy importantes porque su éxito no

puede ser explicado por una teoría de la inteligencia ya que contradicen el dogma de que a mayor inteligencia mayor éxito en su adecuación al entorno. Independientemente de como el individuo procese la información nunca podrá responder adecuadamente al entorno.

## 2.6. Alternativas a la inteligencia

A finales de los años 50, Lee J. Cronbach propuso que la psicología cognitiva y la psicometría tenían que combinarse para entender las diferencias individuales [112]. Sin embargo, su propuesta no empezó a ser desarrollada por la comunidad científica hasta 20 años después de cuando se lanzó. La investigación siguiendo la propuesta de Cronbach ha buscado intentar explicar los conceptos de las teorías psicométricas de la inteligencia se ha estado desarrollando [547]. Los resultados de estas investigaciones han llevado a encontrar relaciones entre las diferencias individuales y procesos cognitivos de bajo y alto nivel. Ese hecho, junto al fracaso continuado en el intento por explicar la inteligencia como propiedad causante de los comportamientos inteligentes, ha provocado que se mire hacia los procesos cognitivos para proponer que la causa de los comportamientos complejos de los seres humanos es la cognición. Así, algunos investigadores han sustituido durante la última década la noción de inteligencia por la de cognición, comportamiento inteligente por comportamiento cognitivo y agente inteligente por agente cognitivo [507, 604]. Por un lado, esto tiene la ventaja de que no es necesario discutir con la definición de inteligencia de la psicometría, y que existen muchos procesos cognitivos muy estudiados como la percepción, categorización, memoria y aprendizaje. Además, la idea de cognición encaja perfectamente con la investigación en IA [507]. Pero, por otro lado, la situación con la cognición no es diferente de la que existe con la inteligencia. No hay una definición de qué es cognitivo y qué no [522, 9, 8] y no son pocos los debates abiertos sobre la cognición [139, 602, 129]. Fred Kizjer resalta la paradójica situación sobre cognición:

“ Interestingly, if cognition is literally held to be equivalent to internal information processing as stated here, any computer would be a cognitive system. Most researches and philosophers active within cognitive science are unwilling to accept this conclusion and additional requirements are standardly used to bar this possibility. ” [261]pag.3.

Por lo tanto, este cambio no mejora la situación, porque tampoco se sabe que es la cognición ni hay una definición científica de qué es.

Es necesario mencionar la propuesta de Frederick Toates de que el comportamiento de los animales podría ser explicado por una interacción entre procesos cognitivos y conductistas de estímulo-respuesta [570]. Aunque esta propuesta tiene numerosas evidencias que la apoyan y ha alcanzado bastante repercusión, hay dos cuestiones a tener

en cuenta. La primera es que no hay una definición de que es cognitivo. La segunda es que la cognición no puede explicar todos los comportamientos porque en la revisión que se ha hecho de los descubrimientos de los comportamientos se ha citado los descubrimientos sobre los comportamientos aleatorios y los ritmos circadianos. En ninguno de estos comportamientos se está procesando información por lo que el concepto de cognición no puede servir para unificar los comportamientos de sistemas biológicos y no biológicos autónomos.

Una alternativa que se ha propuesto para sustituir el paradigma de la inteligencia es la de crear un paradigma basado en las características que poseen los sistemas que están vivos [633]. Pero esto implicaría reconstruir el proceso evolutivo que llevó a la aparición de organismos capaces de comportamientos complejos.

Otra propuesta es la de Jon Doyle de extender la mecánica clásica para explicar la mente humana[128]. Su propuesta intenta describir cómo trabaja la mente y su interacción con el cuerpo usando el marco conceptual de la mecánica clásica. Así, divide los elementos que pueden cambiar en dos tipos: masa y posición. Los elementos de masa son cambiados por aprendizaje y los de posición por razonamiento. Su propuesta es interesante porque permite una unificación de los sistemas bajo las mismas leyes, pero adolece de muchos problemas para explicar los comportamientos[109].

Pero junto a la psicología cognitiva, la línea de investigación que más se ha desarrollado en el intento por formular una teoría que explique el comportamiento de los seres humanos y animales es la teoría de la decisión. La teoría de la decisión se basa principalmente en el modelo de agente racional y la teoría de juegos [184]. Los resultados de la teoría de la decisión permiten entender por qué individuos eligen unas opciones y no otras. El hecho que hay que notar es que la teoría de las decisiones puede ser justificada desde la teoría de la evolución [348, 227]. Un sistema de toma de decisiones no podría conservarse en una población si conduce a la extinción o si existe otro que permite un mayor éxito reproductivo. La definición de agente racional para explicar la toma de decisiones es correcta cuando el agente racional es el mejor biológicamente adaptado al entorno. A pesar de toda la capacidad explicativa de la teoría de decisiones esta se circunscribe solo a la elección entre las distintas opciones. La teoría de la decisión no se ocupa de cómo las opciones son generadas por un organismo, ni cómo los organismos son capaces de transmitirse información entre ellos, ni de como almacenan esa información, ni de cómo recuperar información relevante para toma de decisiones de toda la información almacenada.

## 2.7. Conclusiones del análisis

Tras el análisis realizado en este capítulo, se obtienen las siguientes conclusiones:

- Es razonable buscar alternativas al concepto de inteligencia para explicar los comportamientos inteligentes. Se ha visto que después de tanto esfuerzo, el concepto

no ha sido útil para llegar a nuevos descubrimientos ni desarrollar nuevas tecnologías.

- El concepto de procesamiento de información es insuficiente para formular una teoría que unifique los comportamientos porque existen comportamientos en los que no se procesa información del entorno.
  
- La teoría debe de contemplar las diferencias individuales como las diferencias entre especies.
  
- La unificación de los comportamientos tiene que contener los comportamientos que generen sistemas no biológicos autónomos.
  
- La teoría que se formule para unificar los comportamientos debería hacer uso de la teoría de la evolución y el principio de selección natural. La teoría de la evolución es la teoría científica más sólida que existe y las predicciones de una teoría que unifique los comportamientos deben coincidir con las de la teoría de la evolución. Por ello, lo más sencillo es que la teoría de la evolución este incorporada en la teoría para unificar los comportamientos.

## Capítulo 3

# Evolución y comportamientos

En el capítulo anterior se ha mencionado brevemente que Darwin encontró que el mecanismo de selección natural podía explicar la aparición de los comportamientos instintivos y que consideraba que los de los seres humanos también habían surgido por la selección natural. Darwin escribía:

*“He who wishes to decide whether man is the modified descendant of some pre-existing form, would probably first enquire whether man varies, however slightly, in bodily structure and in mental faculties; and if so, whether the variations are transmitted to his offspring in accordance with the laws which prevail with the lower animals. Again, are the variations the result, as far as our ignorance permits us to judge, of the same general causes, and are they governed by the same general laws, as in the case of other organisms; for instance, by correlation, the inherited effects of use and disuse, etc.?”*  
[115]p.5

En los trabajos de Darwin, el comportamiento fue considerado un rasgo que podía ser explicado por el proceso de selección natural como ya se ha mencionado; pero también reconoció que podía ser un elemento activo como dejó claro al enunciar el principio de selección sexual[115]. Siguiendo esa segunda línea, James Mark Baldwin propuso a finales del siglo XIX que el comportamiento era un factor que afectaba a la evolución [32]. Durante el siglo XX surgieron dos opiniones contrarias sobre cuál era el papel activo del comportamiento en la evolución. Una opinión fue que el comportamiento es una fuerza que produce que aparezca evolución [380, 350]. La otra opinión fue que el comportamiento inhibía el proceso de evolución [59, 465]. Diferentes investigaciones han mostrado que existen comportamientos capaces de desencadenar evolución [287] y otros que son capaces de inhibir evolución [236]. Así, actualmente la opinión más extendida es que no existe un único efecto del comportamiento, sino que el efecto depende del comportamiento concreto que produzca el organismo [132]. Por ejemplo,

un comportamiento exploratorio puede llevar a un organismo a cambiar a un entorno diferente con diferentes presiones evolutivas lo que causará un proceso evolutivo para adaptarse a las nuevas presiones. Por otro lado, un comportamiento migratorio hace que los organismos eviten el cambio de condiciones climáticas y de recursos, lo que causa que no se produzca un proceso evolutivo de adaptación para hacer frente a las condiciones climáticas de los inviernos. Estos ejemplos ponen de relieve que los comportamientos son elementos importantes en la teoría de la evolución. Por ello, una teoría de unificación de los comportamientos tiene que abordar su papel en los procesos evolutivos. Así, si se busca formular una teoría de unificación de los comportamientos es necesario conocer los descubrimientos hechos desde el punto de vista evolutivo de los comportamientos. Así, este capítulo se centrará en abordar el comportamiento desde el punto de vista de la teoría de la evolución.

La evolución y el papel del comportamiento en la evolución son temas demasiado extensos para ser abordados en detalle en un solo capítulo. Por lo tanto, el objetivo de este capítulo es hacer una revisión de las cuestiones más destacadas de la teoría de la evolución relacionados con los comportamientos e introducir algunos conceptos que serán usados posteriormente. Así, la organización del capítulo es la siguiente. La primera sección hace un breve resumen de conceptos y descubrimientos sobre la herencia y desarrollo de organismos biológicos. La segunda sección aborda la formulación abstracta de la teoría de evolución mediante la selección natural de Darwin. La tercera sección trata el tema de las bases genéticas del comportamiento. La cuarta aborda el fenómeno de la plasticidad fenotípica. La quinta aborda cómo el papel del comportamiento en la formulación de la síntesis moderna y de la síntesis extendida de la teoría de la evolución. La última sección presenta las conclusiones de la revisión que se ha realizado.

### **3.1. Herencia y desarrollo en organismos biológicos**

Una cuestión de estudio de la biología son los procesos ontogénicos de los organismos. Un proceso ontogénico es el desarrollo de un organismo, desde la fecundación del cigoto durante reproducción sexual hasta su senescencia, pasando por la forma adulta. Los primeros resultados sobre los mecanismos moleculares de cómo los organismos biológicos se desarrollan se remontan a los trabajos de Gregor Mendel. Mendel dedujo la existencia de una unidad que se transmitía de padres a hijos que determinaba la función y estructura del organismo [363]. En 1909 Wilhelm Johannsen acuñó la palabra gen para referirse a la unidad física y funcional de la herencia biológica. Dos años más tarde escribió un magistral artículo defendiendo la importancia de los conceptos de fenotipo y el genotipo, aunque para él, tenían inicialmente un significado poblacional y no individual [247]. El significado de esos términos en estos momentos es el siguiente. El genotipo de un organismo es toda la información contenida en los cromosomas, y

el fenotipo hace referencia a la información que se ha llegado a expresar del genotipo. En 1941, Edward Lawrie Tatum y George Wells Beadle demostraron que los genes codifican proteínas [39]. En 1942, Conrad Hal Waddington resalta la importancia del estudio de las interacciones entre genes y ambiente que se producen en los organismos y acuñó el término epigenética para referirse ello [598]. Sin embargo, aun pasará mucho tiempo hasta que la epigenética sea considerada un importante elemento a estudiar. En 1943 Oswald Theodore Avery, Colin McLeod y Maclyn McCarty demostraron que el ADN es el material genético [28]. En 1943 Joachim Hämmerling determinó que la ontogénesis está controlada por el núcleo utilizando dos especies distintas de un alga verde: la acetabularia mediterránea que tiene una corola lisa y la acetabularia crenulata que tiene una corola irregular. Hammerling cortó el pie de las algas, extrajo el núcleo de la acetabularia mediterránea y la reemplazó en la acetabularia crenulata y viceversa. De esta manera, el pie que pertenecía a la variedad mediterránea generó una corola crenulata. Es decir, el reemplazo del núcleo provocó una variación en el fenotipo del alga.

Un elemento diferente del genoma surgió en la explicación de los procesos ontogénicos cuando en 1952 Turing escribió un artículo explicando cómo a partir de un estado homogéneo pueden surgir formas no homogéneas mediante patrones de ondas [586]. Ese artículo ha sido la base para la teoría del desarrollo de patrones biológicos [183]. En ella, los procesos de formación de los organismos multicelulares se explican a partir de la combinación de la información almacenada en los genes con los patrones de ondas químicas.

En 1953 James Watson y Francis Crick descubrieron la estructura del ADN [607], lo que permitió se comenzara a estudiar en profundidad el proceso de traducción en las proteínas. En base a los descubrimientos que se hicieron sobre la estructura del ADN, George Gamow postuló que el código genético estaría formado por tripletes de bases nitrogenadas y que a partir de estas se formarían los 20 aminoácidos esenciales para la vida [170]. En 1964 Marshall W. Nirenberg y Philip Leder dilucidaron mediante experimentos los tripletes del código genético y se descifraron los codones que permanecían sin determinar en el código genético. En 1978, Edward B. Lewis descubrió los genes homeóticos que regulan el desarrollo embrionario en la *Drosophila* [301]. Los genes homeóticos codifican proteínas que actúan como factores de transcripción que afectan mediante rutas genéticas regulatorias a otros genes que dirigen el desarrollo del organismo [225]. En 1980, Christiane Nüsslein-Volhard and Eric Wieschaus descubrieron los genes gap [431]. Los genes gap codifican factores de transcripción en un embrión en desarrollo, controlando su subdivisión en partes anterior, media y posterior.

Un importante cambio en la concepción de los procesos ontogénicos se dio tras la finalización del proyecto Genoma Humano, ya que quedó claro que el fenotipo no se podría explicar únicamente a partir de la información de los genes. Así, la epigenética comenzó a ser un campo de un enorme interés. La epigenética busca conocer los mecanismos que controlan la expresión o silenciamiento de los genes. Se sabe que la

regulación epigenética se puede dar por cambios en la conformación de la cromatina. Según el grado de condensación de la cromatina un gen es activado o silenciado. Si encuentra en un alto grado de condensación, los elementos de transcripción no pueden acceder a dicha región del ADN y, por lo tanto, el gen es silenciado. Si la cromatina no se encuentra condensada, ocurre lo contrario. Una reciente investigación ha mostrado la existencia de herencia epigenética [145].

## 3.2. La formulación general de la selección natural

Tras muchas reflexiones, Darwin propuso la existencia de un proceso de evolución en la naturaleza y que ese proceso estaba dirigido por el principio de la selección natural [114]. El principio de evolución mediante la selección natural consiste en que si una población de organismos que tienen variaciones entre ellos y algunas de las variaciones dejan más descendencia que otras, y los descendientes se asemejan a los progenitores, entonces la composición de la población cambiará según avance el tiempo. Existen diferentes marcos para formular el principio de evolución mediante selección natural de Darwin. Uno de los más usados es el de Richard Charles Lewontin que formuló en el artículo *Units of selection* [302] que hoy es un clásico. El marco de Lewontin consta de tres principios: variación fenotípica, diferente adecuación biológica, y heredabilidad. Esos tres principios se pueden definir de la siguiente manera:

- El principio de variación fenotípica: Los individuos de una población presentan diferentes características en sus distintos rasgos.
- El principio de diferente adecuación biológica: Los individuos de la población tienen diferentes tasas de supervivencia y reproducción.
- El principio de heredabilidad: Hay una semejanza entre el fenotipo de los progenitores y su descendencia directa.

En el caso del principio de heredabilidad se ha realizado una pequeña variación respecto a cómo lo enunció Lewontin porque como ya se ha mencionado [433], la formulación de Lewontin se basa en el teorema fundamental de Fisher donde el significado de evolución es el cambio en la media de la adecuación biológica; pero aquí el significado de evolución que interesa es el de variación en el fenotipo.

En toda población donde sus elementos cumplan esos tres principios aparecerá un proceso de evolución por selección natural. A las entidades de una población que cumplen esos tres principios, Lewontin las denominó *unidades de evolución*. Es importante notar que, aunque la teoría de la evolución mediante el principio de selección natural que formuló Darwin fue para explicar la evolución en sistemas biológicos y basada en los hechos que observa en la naturaleza, la teoría no implica que las entidades tengan

que ser sistemas biológicos. Así, por ejemplo, la selección natural es la base del campo de algoritmos genéticos [230]. Incluso, se ha propuesto que las mismas leyes se cumplen en la evolución de poblaciones de sistemas no biológicos y de sistemas biológicos [179, 441].

Es importante mencionar el hecho de que la adecuación biológica es un valor siempre relativo a un entorno debido a que los rasgos que convierten a un organismo en el mejor adaptado en un nicho le hacen no estar adaptado en otros, ya que surgen incompatibilidades.

La teoría de la evolución ha tenido tres grandes formulaciones. La primera fue la formulación que hizo Darwin. La segunda la denominada síntesis moderna [238], y la tercera la síntesis extendida [461]. En ellas el mecanismo de selección natural persiste como lo planteó Darwin.

### 3.3. Las bases genéticas del comportamiento

En la sección anterior se ha visto que uno de los requisitos para que exista un proceso de selección natural sobre un rasgo es la necesidad de que haya heredabilidad del carácter de un rasgo. Por lo tanto, en el caso del comportamiento, un comportamiento solo tendrá efectos evolutivos en las unidades de selección, ya sean produciendo evolución o inhibiéndola, cuando el comportamiento sea transmitido a otros individuos.

En los sistemas biológicos una de las maneras de transmitir comportamientos es mediante la transmisión de material genético de los progenitores a sus descendientes. La discusión de la importancia de la genética en los comportamientos es el famoso debate de ¿nace o se hace? Sin entrar en ese debate, lo seguro es que al menos parcialmente el comportamiento viene determinado genéticamente. Por lo tanto, es necesario mencionar este factor. Un importante elemento en la genética del comportamiento es el tipo de arquitectura genética que tiene el organismo. La arquitectura genética hace referencia a la relación entre genotipo y fenotipo. Se distinguen dos tipos de arquitecturas: genes individuales y múltiples genes. Si el comportamiento viene determinado por múltiples genes que se encuentran en locus<sup>1</sup> distintos, es más fácil que haya una evolución del comportamiento; pero si el comportamiento viene determinado por un solo locus entonces se alcanzará antes un comportamiento óptimo en el caso de que exista. Un caso de comportamiento controlado por un locus fue descubierto en la *Drosophila melanogaster*. En investigaciones llevadas a cabo se vio que la *Drosophila* tiene un comportamiento de explorador o de niñera dependiendo de las mutaciones de un gen [439]. En los últimos años se ha logrado encontrar los genes responsables de comportamientos de algunos organismos concretos, pero ese hecho es a día de hoy una excepción. Hasta el momento, el mayor número de investigaciones realizadas se han

---

<sup>1</sup>Un locus es una posición fija en un cromosoma

centrado en los genes responsables de los comportamientos de elección de pareja y de los de comportamientos sociales. Entre los descubrimientos que se han hecho en relación con los comportamientos se ha encontrado la existencia de acoplamiento genético en la *Drosophila melanogaster* [337]. El acoplamiento genético implica que rasgos de machos y hembras involucrados en la selección de pareja dependen del mismo locus. Concretamente, se descubrió que la producción de feromonas para las señales sexuales y su detección dependían del mismo locus. Hasta qué punto los rasgos dependientes entre machos y hembras están acoplados genéticamente en todas las especies es una cuestión abierta. Otro descubrimiento importante ha sido como un gen neuroendocrino controla el comportamiento monógamo en los topos de las praderas [309, 437]. Un último descubrimiento por citar es el comportamiento de las abejas en las colmenas. Las abejas cambian la tarea que realizan en la colmena según envejecen [520]. Un estudio genético de las abejas mostró que la transición de comportamiento relacionada con la edad de las abejas está asociada con el incremento de la expresión de un gen [52]. Otra cuestión es la pleiotropía en la evolución del comportamiento. La pleiotropía es el fenómeno por el cual un solo gen es responsable de efectos en rasgos distintos y no relacionados. Algunos investigadores han sugerido que la pleiotropía que afecta al comportamiento podría ser un hecho común [194], pero hasta ahora solo se tienen algunos ejemplos en los que ocurre [364].

El avance en los últimos años en las técnicas genéticas está mostrando un enorme potencial para investigar las bases genéticas del comportamiento [14, 241]. Así que se puede decir que la investigación está tan solo comenzando.

### 3.4. La plasticidad fenotípica

La plasticidad fenotípica es la propiedad que posee un genotipo para producir diferentes fenotipos en respuesta a condiciones de entorno diferentes [553]. La plasticidad fenotípica es una característica por sí misma controlada genéticamente y que puede evolucionar independientemente de otros aspectos del fenotipo. Además, hay que tener en cuenta que la plasticidad no es una característica de todo el fenotipo; se necesita estudiar en referencia a unos rasgos y entornos concretos. Desde el punto de vista evolutivo la plasticidad fenotípica ha despertado interés en los últimos años.

Dos ejemplos clásicos de plasticidad fenotípica son: los ocelos en las alas de algunas mariposas [493] y la respuesta que presentan diversas plantas para evitar la sombra [462].

Las mariposas *Bicyclus anynana* experimentan un cambio drástico regular en su entorno que cambia completamente de la estación seca a la estación húmeda. Durante la estación seca apenas hay recursos por lo que las mariposas lo único que pueden intentar es sobrevivir. Así, las mariposas apenas se mueven durante la estación seca y en su alas los ocelos difícilmente pueden apreciarse. Pero durante la estación húmeda,

las mariposas están muy activas y su estrategia de defensa cambia haciendo que los ocelos se vean perfectamente.

El caso de plasticidad fenotípica de las plantas es debido a la competición que aparece entre ellas por la luz solar. Cuando una planta detecta que está siendo tapada por otras plantas, hay dos posibles estrategias que se pueden realizar. Una estrategia es dejar de desarrollar ramas laterales para acelerar su crecimiento vertical para ganar a sus competidoras y acceder a la luz del sol directamente. La segunda estrategia es acelerar su ciclo de vida para florecer rápidamente y evitar a los competidores.

Una cuestión importante es que la plasticidad fenotípica es una característica que tiene un coste para el organismo [119]. Este hecho hace que la adecuación biológica del organismo con plasticidad fenotípica pudiera ser inferior a la de otro organismo en sin plasticidad fenotípica en un entorno sin cambios a los que sea necesario responder. Esto ha llevado a considerar el aspecto adaptativo de la plasticidad fenotípica. Múltiples preguntas relacionadas con cómo la plasticidad fenotípica es dirigida por la selección natural y cómo la evolución es afectada por la plasticidad fenotípica han surgido [629].

Una última cuestión con referencia a la plasticidad fenotípica es resaltar que aunque mucha de la investigación sobre plasticidad fenotípica se centra en rasgos morfológicos y fisiológicos, el comportamiento es un rasgo de los organismos y que por lo tanto también puede presentar plasticidad fenotípica [178]. De hecho, se conocen múltiples comportamientos que varían en respuesta al entorno en el que esté el organismo y que dependen del genotipo [482].

### 3.5. Evolución y comportamiento

Al inicio del capítulo se mencionó el debate que ha existido sobre qué papel ha llevado a cabo el comportamiento en la evolución organismos. Aparte de la discusión de si el comportamiento dirige o inhibe la evolución, existe un problema para abordar la concepción del comportamiento en la formulación de la síntesis moderna de la teoría de la evolución. La síntesis moderna relaciona la genética mendeliana, la selección natural y la teoría cromosómica de la herencia [238]. Ese problema fue expuesto por Richard Lewontin mediante un sistema de dos ecuaciones diferenciales [303].

$$\frac{d(O)}{dt} = f(O, E) \quad (3.1)$$

$$\frac{d(E)}{dt} = g(E) \quad (3.2)$$

En la primera ecuación del sistema el término  $\frac{d(O)}{dt}$  representa el cambio evolutivo del organismo,  $O$  representa el estado del organismo, y  $E$  representa el estado del entorno. La ecuación expresa que el cambio evolutivo del organismo es función del estado

del entorno y del estado del organismo. La segunda ecuación expresa que el cambio del estado del entorno,  $\frac{d(E)}{dt}$ , es función únicamente del estado en el que se encuentra el entorno. Según ese sistema de ecuaciones la evolución del entorno es independiente del organismo. Por lo tanto, la síntesis moderna no permite reconocer al comportamiento como un elemento activo, ya que solo explica que el comportamiento es un rasgo evolutivo del fenotipo que afecta al éxito reproductivo de los organismos y por lo tanto afecta a la selección de un determinado genotipo. Así, Lewontin propuso que la teoría de la evolución debería de ser formula con el siguiente sistema de ecuaciones:

$$\frac{d(O)}{dt} = f(O, E) \quad (3.3)$$

$$\frac{d(E)}{dt} = g(O, E) \quad (3.4)$$

En este nuevo sistema de ecuaciones, el cambio del entorno,  $\frac{d(E)}{dt}$ , ya no es función únicamente del estado del entorno, sino que depende también del estado del organismo. El problema de representar el comportamiento como un elemento activo en la evolución ha llevado a formular la teoría de la construcción del nicho [432, 284]. La teoría de la construcción del nicho afirma que la construcción del nicho debería ser reconocida como un proceso que afecta a la evolución de los organismos. La teoría de la construcción del nicho es una de las extensiones que se ha añadido en la síntesis extendida de la teoría de la evolución [461]. Así, en la comprensión que se tiene actualmente de la evolución, el comportamiento es un elemento activo que es capaz de promover o inhibir el proceso evolutivo.

### 3.6. Conclusiones de la revisión

Tras las cuestiones abordadas y las discusiones realizadas, se pueden resaltar los siguientes puntos:

- El proceso ontogénico consiste en un conjunto de mecanismos genéticos y moleculares que transforman parte de la información almacenada en el genotipo en el fenotipo.
- El principio de selección natural se cumple en todo grupo de entidades (biológicas o no biológicas) en el que se cumpla: variación fenotípica, diferente adecuación al entorno, y heredabilidad.
- El comportamiento puede producir o inhibir cambio evolutivo. Así, el que los organismos de una población se comporten de una manera u otra tendrá efectos en la evolución de la población.

- En los sistemas biológicos, el comportamiento viene determinado, al menos parcialmente, por los genes.
- La plasticidad fenotípica es un rasgo que permite a los organismos responder al estado del entorno.



## Capítulo 4

# Teoría general del exocomportamiento

En el capítulo 2 se ha concluido que es necesario investigar alternativas a los conceptos de inteligencia y cognición para desarrollar una teoría que unifique los comportamientos de sistemas biológicos y no-biológicos autónomos y dé una explicación científica de ellos. Este capítulo presenta la teoría general del exocomportamiento (TGE), una propuesta de unificación de los comportamientos que se basa en una concepción física del comportamiento evitando los conceptos del paradigma mentalista.

En la revisión que se ha hecho en el capítulo 2 de las propuestas para explicar los comportamientos, la mayoría se han formulado bajo la hipótesis de que los comportamientos de los seres humanos son un conjunto de fenómenos diferente a los que realizan el resto de organismos. Ese planteamiento de partida impide que pueda existir una unificación de los comportamientos con esas teorías. En cambio, la TGE parte de la hipótesis de que todos los comportamientos pertenecen a un mismo conjunto de fenómenos de la naturaleza, y, por lo tanto, tienen que tener la misma causa. La formulación de la TGE se ha desarrollado usando una variación del método de Isaac Newton para razonar sobre la naturaleza. Las ideas fundamentales de la TGE fueron formuladas en trabajos previos a los de esta tesis doctoral [372, 368], pero en esta tesis se presenta una formulación que varía de las anteriores. La nueva formulación resuelve alguna cuestión problemática de anteriores formulaciones, cumpliéndose el objetivo de la tesis que se presenta de investigar y desarrollar los fundamentos de la TGE.

Primeramente, se va a mencionar brevemente alguna terminología sobre la explicación científica. Después se va a presentar la variante del método de Newton que se ha ideado para realizar razonamientos inductivos con los que formular la TGE. A continuación, se comenzará con la formulación de la TGE. El primer paso será definir un conjunto de fenómenos que incluirá todos los comportamientos de sistemas biológicos y sistemas no-biológicos autónomos que se conocen. Este conjunto es denominado, exocomportamientos. Una vez se definan los exocomportamientos, siguiendo el método

de Newton se le asignará una causa al conjunto de fenómenos, y se buscará una interpretación que explique qué es esa causa. La causa en la TGE se denomina *fasa*. En ese punto se expondrá qué leyes y postulados posee la TGE para explicar los exocomportamientos en base a la *fasa*. Tras eso, se comentará brevemente los cambios realizados en la formulación que se ha presentado de la TGE respecto de las anteriores formulaciones. Finalmente, se hará un análisis de la investigación presentada en este capítulo.

## 4.1. La explicación científica

El primer trabajo minucioso sobre qué es una explicación científica fue artículo de Carl Hempel y Paul Oppenheim de 1948 [219]. Su trabajo desencadenó una profunda discusión sobre la cuestión. Desde entonces la filosofía de la ciencia ha buscado lograr precisar qué es una explicación científica [498]. Este problema queda fuera de los objetivos de esta tesis. Sin embargo, la terminología que se usa en filosofía de la ciencia para abordar la explicación científica es útil para explicar la formulación de la TGE. Los siguientes cuatro términos serán usados en este capítulo a menudo:

- *Explanandum* (o *Explicandum*): definición informal de aquello que se busca explicar.
- *Explanatum* (o *Explicatum*): definición formal de aquello que se busca explicar.
- *Explanans* (o *Explicans*): aquello que proporciona la explicación.

El punto de partida de la investigación que presenta este capítulo era el *explanandum* que forman los comportamientos. Así, el objetivo ha sido lograr primero un *explanatum* para el *explanandum*, y luego un *explanans* para el *explanatum*. Un requisito que se pide a un *explanatum* es que tenga similaridad con el *explanandum* [498]. Sin embargo, el grado de similaridad no tiene por qué ser muy alto, ya que al final dependerá de si el *explanandum* describe prácticamente completo a un conjunto de fenómenos iguales de la naturaleza o si es demasiado heterogéneo.

## 4.2. Razonando sobre sistemas naturales

En la sección 1.2 se mencionó que una de las hipótesis de trabajo para la investigación es el fisicalismo de los comportamientos. Esa hipótesis determinará la naturaleza de los conceptos que se formularán para explicar los comportamientos; pero es necesario también un método que permita formular proposiciones sobre los fenómenos que se intentan explicar. Así, a continuación se presentará un método para razonar sobre fenómenos naturales que se aparecen en sistemas físicos con frontera. El método que

se presenta es una variación de las reglas de razonamiento Newton. Newton usó su método para realizar razonamientos inductivos que le permitieron formular su teoría para unificar la mecánica celeste y la terrestre. El método de Newton se aplica a las propiedades de la materia. Sin embargo, el fenómeno del comportamiento solo aparece en sistemas físicos con frontera. Esto obliga a modificar el método de Newton.

#### 4.2.1. Las reglas de razonamiento de Newton

El primero en lograr una teoría que unificó dos conjuntos de fenómenos naturales fue Newton. Él presentó la teoría de unificación en su *Philosophiae Naturalis Principia Mathematica* [421]. En ese manuscrito, además de su teoría, presentó un método para razonar sobre la naturaleza que le permitió hacer razonamientos inductivos para justificar las afirmaciones de su teoría. Su método estaba basado en el método de Galileo. Pese a las discusiones que ha habido sobre su método, este puede concluirse correcto [541]. El método que Newton propuso para razonar sobre la naturaleza tiene las siguiente cuatro reglas:

**Regla 1** “*We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances.*”[422]p.378.

**Regla 2** “*Therefore to the same natural effects we must, as far as possible, assign the same causes.*”[422]p.378.

**Regla 3** “*The qualities of bodies, which admit neither intensification nor remission of degrees, and which are found to belong to all bodies within the reach of our experiments, are to be esteemed the universal qualities of all bodies whatsoever.*” [422]p.378.

**Regla 4** “*In experimental philosophy we are to look upon propositions inferred by general induction from phenomena as accurately or very nearly true, notwithstanding any contrary hypothesis that may be imagined, till such time as other phenomena occur, by which they may either be made more accurate, or liable to exceptions.*”[422]p.400.

Este conjunto de reglas permitieron a Newton concluir que la gravitación era un fenómeno universal de la materia causado por la propiedad de la masa.

### 4.2.2. Un método para razonar sobre sistemas físicos

El objetivo principal de la investigación que se ha llevado a cabo ha sido la unificación de los comportamientos de sistemas biológicos y no biológicos autónomos. Dado que el método de Newton fue el método empleado por el propio Newton para razonar la primera teoría de la unificación de la ciencia, sin duda es un esquema de razonamiento plausible para desarrollar una teoría. Sin embargo, existen ciertas cuestiones que requieren modificar su sistema de razonamiento para usarlo sobre los comportamientos. Concretamente, la cuestión principal es que los fenómenos de los comportamientos no son encontrados en todos los cuerpos. A diferencia de los efectos gravitatorios que son producidos por cualquier cuerpo, en el caso de los comportamientos solo ciertos sistemas presentan ese fenómeno.

Para solventar el problema, se propone una modificación del método de Newton para razonar sobre clases de sistemas físicos que ayude a formular una teoría que explique los comportamientos. El método que se propone es el siguiente:

**Regla 1** *Admitir que no hay más causas de los objetos naturales que las que son verdaderas y suficiente para explicar sus aspectos.*

**Regla 2** *Por lo tanto, a los mismos efectos naturales, en la medida de lo posible, se les asigna las mismas causas.*

Las dos primeras reglas son las mismas que las del método de Newton.

**Regla 3** *Las cualidades de objetos, que no admiten intensificación ni remisión, sin que el sistema sea modificado, y que son encontradas en todos los objetos de una clase dentro del alcance de nuestros experimentos, deben ser estimadas como cualidades universales de cualquier objeto de esa clase.*

La regla tres se modifica respecto a la de Newton porque la regla 3 de Newton tiene como dominio cualquier objeto de la naturaleza, ya que los fenómenos sobre los que él buscaba razonar aparecen en todos los objetos con los que no se podía realizar experimentos y necesita hacer una inducción que afecte a cualquier objeto

del universo. Sin embargo, los comportamientos son fenómenos que no aparecen en cualquier objeto de la naturaleza sino solo en una clase de objetos. Por lo tanto, en el caso de los comportamientos, las conclusiones de un razonamiento inductivo tienen que circunscribirse a una clase de objetos. La importancia de tener un dominio cerrado al que se circunscriban la formulación de una teoría está relacionado con los procesos de deducción que se llevan a cabo para explicar fenómenos una vez se tiene la formulación de la teoría. Si las proposiciones de las que se parte en un razonamiento deductivo son verdad en un dominio cerrado, las conclusiones serán verdad en ese dominio cerrado si las reglas de deducción son correctas.

Un punto importante que se debe notar sobre la regla 3 es que distingue cualidades que pertenecen al estado del sistema, de cualidades que conforman la identidad del sistema. Mientras las cualidades del estado del sistema pueden cambiar sin que se altere la identidad del sistema las que conforman el sistema no pueden cambiar sin que el sistema pierda su identidad.

**Regla 4** *En ciencia, hemos de ver las proposiciones obtenidas por inducción, a partir de los fenómenos como completamente, o prácticamente, verdad, sin importar cualesquiera hipótesis contrarias que se puedan imaginar, hasta que otro fenómeno haga tales proposiciones más exactas o sujetas a excepciones.*

En esta regla solo se ha cambiado el término filosofía experimental por el término ciencia, ya que el uso del término de filosofía experimental que hizo Newton estuvo causado por sus debates con cartesianos y leibnizianos [527].

### 4.3. Exocomportamientos: el conjunto de fenómenos

Formular un *explanans* de un conjunto de fenómenos requiere un *explanatum* de ese conjunto. Recuérdese que un *explanatum* es una definición precisa de un *explanandum*. Puesto que el *explanandum* que se aborda son fenómenos físicos, una definición precisa de un conjunto de fenómenos físicos implica el uso de parámetros físicos que pueden ser observados y medidos para decidir si un fenómeno pertenece o no al conjunto de fenómenos definido. En la bibliografía no se suele dar una definición de lo que es el comportamiento como fenómeno, y si se da, se trata de una definición más bien intuitiva, alejada de una definición formal y precisa. Por lo tanto, se tiene un *explanandum*, y es necesario lograr un *explanatum*.

A pesar del tiempo transcurrido desde el nacimiento de las ciencias del comportamiento, hasta ahora solo se tiene un *explanandum*; aunque sí existen dos métodos generales para la descripción de comportamientos de organismos concretos en entornos concretos [235]. Uno de los métodos describe un comportamiento como una secuencia de los efectos que el organismo produce en el entorno. El otro método lo describe como la secuencia de actividades que realiza el organismo con su cuerpo. Usar un lenguaje anatómico condiciona al *explanatum* a la anatomía de organismos concretos. Usar un lenguaje sobre los efectos en el entorno condiciona también a unos entornos específicos. Por lo tanto, aunque los dos métodos son útiles para estudiar el comportamiento específico de uno o varios organismos, ninguno lo es para formular una definición precisa de un conjunto de fenómenos que abarque un rango tan amplio de comportamientos como el que se pretende. Esto no quiere decir que los métodos de descripción de comportamientos concretos no tengan valor. Recuérdese que el *explanatum* tiene que mantener un alto grado de similitud con el *explanandum*, sino no podría sustituirle. Por lo tanto, el objetivo es obtener una definición precisa que para cada fenómeno que la cumpla pueda además ser descrito por los dos métodos para describir comportamientos citados.

Primeramente, debe notarse que al pretenderse unificar los comportamientos de los organismos biológicos con los de los no biológicos autónomos, el *explanatum* debe abarcar a ambos. Observando los comportamientos de seres vivos y dispositivos automáticos, se encuentra el siguiente listado: comportamientos motores (por ejemplo, los movimientos de animales, robots, o de las hojas de las plantas), comportamientos luminiscentes (ej. luciérnagas), comportamientos de mimetización activos (ej. el camaleón o el pulpo), comportamientos eléctricos (ej. la anguila eléctrica), comportamientos térmicos (ej. los mamíferos o las plantas de loto), comportamientos de secreción (ej., los pulpos, las plantas que emiten terpenos) y comportamientos de alteración del fenotipo (ej. la *escherichia coli*). Todos estos comportamientos parecen diferentes entre ellos, pero dado que por la hipótesis del fisicalismo de los comportamientos se asume que los comportamientos son fenómenos físicos, la definición del *explanatum* de esta investigación debe ser formulado mediante un análisis de estos fenómenos en términos físicos. Así, usando el punto de vista de la física para analizar el listado mencionado, se encuentran características que son comunes a todos ellos. Las características comunes son tres: los fenómenos se presentan en sistemas físicos con fronteras, en los fenómenos interviene un gasto de energía interna para controlar el fenómeno y hay un flujo de energía que produce el fenómeno. Un sistema físico con frontera es una entidad física compuesta únicamente por la frontera y los elementos físicos que quedan dentro de ella. La frontera separa el interior del sistema físico del exterior, y permite que las condiciones del exterior y el interior del sistema sean diferentes. Tanto los seres vivos, como los sistemas no biológicos autónomos, son sistemas físicos con frontera. Sin embargo, no todos los sistemas físicos con frontera presentan los fenómenos que se buscan explicar. Por lo tanto, se ha de precisar en qué sistemas físicos se observa el *explanatum*. Para precisar los sistemas en los que aparece el *explanatum* se debe analizar el papel de la

energía en las dos características comunes que se han citado. La idea es ver en qué casos la energía determina cuando se trata de un fenómeno del explicatum y cuando no. Las características que se pueden extraer de analizar la energía en los comportamientos son las siguientes:

1. **El sistema físico tiene circuitos para suministrar energía que permita que las propiedades dinámicas que son observadas y medibles del sistema cambien.** Por ejemplo, en los animales hay circuitos para suministrar energía a los músculos, en los robots hay circuitos para el suministro de energía a sus motores. Desde una posición fisicalista y por las propias leyes de la física, no puede haber cambios sin que haya un aporte de energía. Por lo tanto, el sistema debe de tener circuitos que suministren la energía para los cambios.
2. **El sistema tiene una fuente de energía disponible para alimentar los circuitos de energía de las propiedades dinámicas observables y medibles del sistema.** La importancia de que esta característica se cumpla es la siguiente. Imagínese un robot que tiene una fuente de energía para alimentar la CPU y otra para alimentar los motores. Se da la situación de que la fuente de energía para alimentar los motores está agotada. Si el robot envía el comando de activar uno de sus motores para moverse el robot no puede moverse porque la fuente de energía está agotada y por lo tanto el robot no se moverá. El comportamiento del robot que se observaría en esta situación es el de que está inmóvil; pero ese comportamiento no es atribuible al robot, ya que, que el sistema no se mueva no es determinado por las decisiones del robot. Por lo tanto, es un fenómeno que no debe pertenecer al conjunto de fenómenos que se busca explicar.
3. **Las fuentes de energía no son reguladas desde el exterior del sistema.** La razón de exigir esta característica es la siguiente. Imagínese un robot en el que la cantidad de energía liberada por las fuentes de energía que alimentan los circuitos de sus motores puede ser modificada desde el exterior. Si ese robot envía el comando para moverse rápidamente; pero desde el exterior se limita la cantidad de energía que libera la fuente hacia sus motores, el resultado sería que se movería lentamente y la orden del robot no tendría ningún efecto. El comportamiento que se observa de moverse lento no puede ser atribuido al robot. Por lo tanto, es un fenómeno que no debe pertenecer al conjunto de fenómenos que se busca explicar.
4. **Los flujos de energía no son regulados desde el exterior.** La exigencia de esta característica es para evitar también que formen parte del *explanatum* fenómenos que no deben serlo. Imagínese que el circuito que suministra energía a los motores de un robot es regulable mediante un potenciómetro que accesible desde el exterior del sistema. Si se aumenta el valor del potenciómetro desde el exterior ocurrirá que, aunque el robot envía el comando para moverse rápido, el

robot se moverá lento debido a que el potenciómetro deja pasar poca energía. Este comportamiento de moverse lento no es atribuible a las decisiones del sistema, por lo tanto, tampoco debe formar parte del conjunto de fenómenos que se quieren explicar.

5. **La energía de los flujos no puede ser dirigida desde fuera del sistema.** Esta característica también evita que formen parte del *explanatum* fenómenos que no deben hacerlo. Imagínese un robot que posee un interruptor seleccionador que intercambia los circuitos de los motores de las ruedas con los de los brazos y que puede ser manejado desde el exterior. Si el robot envía el comando para activar los motores de las ruedas, pero desde el exterior se cambia el circuito y se seleccionan los brazos, entonces el robot moverá sus brazos. Este tipo de comportamiento tampoco puede ser atribuido al robot, ya que él pretendía activar los motores de sus ruedas. Por lo tanto, no debe formar parte del *explanatum*.

Teniendo en cuenta las anteriores características, se definen tres tipos de sistemas físicos con frontera: *sistemas activos*, *sistemas endoregulados* y *sistemas endodirigidos*.

**Definición 1** (Sistema activo). *Se define un sistema activo como un sistema físico con frontera que tiene fuentes y circuitos de energía que suministran energía a propiedades dinámicas que son observables y medibles del sistema.*

**Definición 2** (Sistema endoregulado). *Se define un sistema endoregulado como un sistema físico con frontera cuyos circuitos o fuentes de energía que suministran energía a propiedades dinámicas que son observables y medibles del sistema pueden regularse desde el exterior del sistema.*

**Definición 3** (Sistema endodirigido). *Se define un sistema endodirigido como un sistema físico con frontera en el que la energía de sus circuitos de energía que suministran energía a propiedades dinámicas que son observables y medibles del sistema puede ser desviada de unos a otros desde el exterior del sistema.*

Usando las anteriores definiciones se define el tipo de objetos que presentan los fenómenos que se buscan explicar, *los sistemas exoactivos*.

**Definición 4** (Sistema exoactivo). *Un sistema exoactivo es un sistema que cumple que es activo, no endoregulado ni endodirigido.*

Nótese que la característica de activo permite abarcar tanto sistemas biológicos como sistemas no biológicos autónomos. Las otras dos permiten eliminar sistemas que presentan fenómenos que no formen parte del *explanatum*. Los sistemas exoactivos son un dominio cerrado en el que ocurre el *explanatum*, lo cual es un hecho importante porque en este dominio se puede aplicar la variante del método de Newton que se ha propuesto para hacer la formulación de la teoría de la unificación que se pretende.

Una vez que se ha definido el conjunto de objetos físicos que presentan el fenómeno, debemos pasar a cuál es el fenómeno que se presenta en esos sistemas. En psicología, etología, o psiquiatría, un comportamiento es una secuencia finita de acciones, o de efectos en un entorno, generada por un organismo [342]. La secuencia comienza con la introducción del organismo en un entorno, o introduciendo un nuevo estímulo en su entorno, y que dura mientras el organismo permanece en el nuevo entorno, o mientras el nuevo estímulo es mantenido en el entorno. Por lo tanto, un comportamiento es entendido como una secuencia concreta de acciones que sucede durante el experimento por la necesidad de describir los efectos que aparecen durante los experimentos. Pero esta concepción de los comportamientos tiene un problema. Si un comportamiento comienza con la aparición de un estímulo o por un nuevo entorno, la causa del comportamiento sería el estímulo o el nuevo entorno. Sin embargo, la psicología cognitiva atribuye el comportamiento al procesamiento interno de información. Esto implica una definición completamente diferente, ya que, según la propuesta de la psicología cognitiva, no hay una diversidad de comportamientos sucediéndose en cadena en el sujeto, sino que el comportamiento se va generando según la información que se va procesando. Esta contradicción en la concepción de los comportamientos no es debida a la existencia de nociones enfrentadas, en realidad es simplemente debida a cuestiones metodológicas del método científico. Psicólogos, etólogos, o psiquiatras deben estudiar el comportamiento de sus sujetos registrando y midiendo los efectos que aparecen durante el experimento. Por lo tanto, no hay muchas secuencias de fenómenos sino uno solo, y ese fenómeno es el comportamiento del sistema durante toda su existencia. Además, si se considera la existencia de una secuencia de comportamientos, el *explanatum* debería tener una causa externa, ya que el *explanatum* aparecería con el nuevo entorno o la aparición del estímulo, y desaparecería al desaparecer los desencadenantes. Ese planteamiento no tiene sentido, por lo que el *explanatum* debe durar durante toda la existencia del sistema exoactivo para que la causa resida en el propio sistema exoactivo.

En anteriores formulaciones de la TGE se ha considerado que el *explanatum* consistía la secuencia de los valores de las propiedades dinámicas observables y medibles desde el exterior [372, 368, 369]. Sin embargo, en la investigación que se ha llevado a cabo sobre la formulación de la TGE, se ha llegado a la conclusión de que el *explanatum* no lo pueden conformar los valores de las propiedades dinámicas observables y medibles desde el exterior, y que son necesarios dos cambios. La razón de uno de los cambios es que los valores de las propiedades dinámicas podrían ser el resultado de una interacción entre el sistema exoactivo y otros sistemas. El registrar los valores de las propiedades puede hacer que como consecuencia de las interacciones el comportamiento del sistema exoactivo quede oculto. Por ejemplo, imagínese dos personas de igual fuerza practicando el deporte de echar un pulso en el que sus brazos están inmóviles porque sus fuerzas se anulan. Si para definir su comportamiento se toma la posición de sus brazos determinaríamos una situación muy diferente de si nos fijamos en la energía que está siendo suministrada a los músculos de sus brazos. Otro ejemplo sería un niño

corriendo para cambiar su posición y el mismo niño montado sobre un monopatín que abre su chaqueta durante un día de viento para ser desplazado por el viento. En los dos casos el tomar los valores de las propiedades impide que se pueda distinguir el comportamiento real de otro que es diferente. Sin embargo, si se toma la energía aportada a sus músculos se diferencia perfectamente cada comportamiento. Así, se consideró que en la definición del *explanatum* se encontrarían las cantidades de energía aportadas por el sistema exoactivo para conformar los valores de las propiedades dinámicas que son observables y medibles desde el exterior. Sin embargo, analizando esa conclusión se vio que romper los procesos ontogénicos en función de si la propiedad es observable desde el exterior o el interior es artificial. Así, aunque las propiedades que se consideran en el *explanandum* siempre son exteriores, no existe un motivo para que no se consideren las interiores para el *explanatum*. Un ejemplo de organismo que modifica su metabolismo en función del entorno es la *Escherichia coli* que puede responder al cambio de nutrientes en el medio cambiando las enzimas que usa para digerirlos. Por lo tanto, se debía realizar un segundo cambio que consiste en no diferenciar entre si la propiedad es medible desde el exterior o el interior de la frontera.

Combinando la definición física de los objetos que presentan el *explanatum*, el tomar como duración del fenómeno toda la existencia del sistema exoactivo, que las propiedades pueden ser tanto internas como externas y que los valores de la cantidad de energía aportada a las propiedades dinámicas evitan ambigüedades, se define el *explanatum* de los comportamientos, denominado *exocomportamiento*, de la siguiente manera:

**Definición 5** (Exocomportamiento). *Se define un exocomportamiento como la secuencia de las cantidades de energía que un sistema exoactivo aporta a través de sus circuitos energéticos para conformar los valores de las propiedades dinámicas que son observables y medibles.*

Según la definición dada de un exocomportamiento, se trata de una secuencia formada por las cantidades de energía en los circuitos del sistema que intervienen en la conformación del valor de las propiedades dinámicas que son observables y medibles. Las cantidades de energía que pueden circular por los flujos de energía de un sistema exoactivo es un conjunto finito de valores,  $C$ . Se denotará como  $c_x^{(t,t']}$  a una cantidad de energía  $c_x$  que pertenece a  $C$  y que es aplicada por un sistema exoactivo durante el intervalo de tiempo  $(t, t']$ . Con esta notación, el exocomportamiento de un sistema exoactivo, al que se denotará por  $\varepsilon$ , respecto a alguna de sus características observables y medibles desde la exterior queda definido de la siguiente manera:

$$\varepsilon = \begin{matrix} (0,u] & (u,v] & (v,w] \\ c_x & c_y & c_z, \dots \end{matrix} \quad u < v < w$$

La duración del exocomportamiento de un sistema exoactivo se extiende mientras el sistema tiene energía disponible para mantener los flujos de energía que el mismo determina. En el caso de los seres vivos esto coincide con la existencia del sistema. Esto

cambia en el caso de sistemas no biológicos, ya que la desaparición de su condición de sistema exoactivo por no disponer de carga en sus fuentes de energía no implica la pérdida de su condición de sistema. El sistema podría volver a tener carga en su fuente de alimentación, y por lo tanto, recuperar su condición de sistema exoactivo.

Una vez hecha la definición del *explanatum*, hay que considerar que fenómenos entran en esa definición. Con esa definición entran los comportamientos de sistemas biológicos y de sistemas no biológicos autónomos, pero también entra el desarrollo de un fenotipo en los sistemas biológicos. Nótese que los procesos ontogénicos necesitan un flujo de energía y que provocan el cambio de características morfológicas y funcionales que son observables y medibles. Así, cualquier proceso ontogénico es un exocomportamiento. A la vez, esto implica que el conjunto de los exocomportamientos contiene más elementos de los que se buscaba unificar; pero esto no tiene que ser un problema de la teoría; sino todo lo contrario. Cuantos más conjuntos de fenómenos unifique una teoría más está explicando sobre la naturaleza. El conjunto de exocomportamientos contiene los comportamientos de sistemas biológicos y de sistemas no biológicos autónomos que se buscaba. Así, se cumple con el objetivo que se buscaba. La cuestión importante es si el conjunto de exocomportamientos es realmente un conjunto que existe en la naturaleza o es una definición artificial. La resolución a esa cuestión depende de si se encuentra una causa que es capaz de explicar todos los fenómenos del conjunto. Si no se encontrara esa causa, la definición sería artificial y no se trataría de un auténtico conjunto de fenómenos de la naturaleza.

Una última cuestión sobre los exocomportamientos es que un lector podría preguntarse cómo es posible con la definición de exocomportamiento saber que dos organismos de distintas especies están llevando a cabo el mismo comportamiento, ya que por sus diferencias biológicas las cantidades de energía que genera cada individuo en su exocomportamiento serán diferentes. Sin embargo, para manejar esta cuestión es posible establecer la equivalencia entre exocomportamientos. El modo de establecer la equivalencia entre exocomportamientos de sistemas exoactivos diferentes es mediante los métodos generales existentes para la descripción de comportamientos que ya se han mencionado. Por lo tanto, los métodos que actualmente existen para registrar y comparar comportamientos tienen su cabida e interés en la TGE.

#### 4.4. Fasa: causa única de los exocomportamientos

Con el objetivo de aplicar el método de razonamiento de Newton se ha definido un *explicatum* que consiste en un conjunto de fenómenos naturales que abarca los comportamientos de sistemas biológicos y no biológicos. Para crear un *explanans*, la regla 2 dice que a fenómenos iguales se les debe asignar las mismas causas. Por lo tanto, si el *explanatum* es el conjunto de los exocomportamientos, en base a la citada regla, se debe asignar una misma causa a todos los fenómenos que contiene. Así, la primera

proposición de la TGE es la siguiente afirmación:

**Postulado 1.** *Los exocomportamientos son causados por una propiedad de los sistemas exoactivos, la fasa.*

La elección del nombre de fasa para la causa de los exocomportamientos es motivada por el propósito de que la terminología de la TGE no referencie a ninguna idea preconcebida, y evitar discusiones, o debates, sobre su significado, para centrarse en obtener su descripción solo mediante la investigación de los fenómenos con los que se asocia. Desde el punto de vista de la ciencia, lo que importa es obtener un *explanans* que permitan obtener *explicatum* que la causa explique los fenómenos, así con esta elección se evitan debates filosóficos que no son útiles en la creación del *explicatum* que se busca.

El postulado 1 puede expresarse mediante la siguiente ecuación:

$$\varepsilon(t) = C(\mathcal{F}(t)) \quad (4.1)$$

donde  $C$  son los circuitos y fuentes de energía del sistema exoactivo y  $\mathcal{F}$  denota un valor de fasa. La ecuación 4.1 expresa que la cantidad de energía que tiene en sus circuitos el sistema en el instante  $t$  es determinada por el estado de fasa en ese momento del tiempo. Nótese que el significado del valor de la fasa de un sistema exoactivo es completamente diferente del estado de fasa de un sistema exoactivo. El valor de fasa de un sistema exoactivo está directamente ligado a la identidad del sistema y se denota por  $\mathcal{F}$ , mientras que el estado puede cambiar sin que cambie la identidad del sistema, y se denota por  $\mathcal{F}(t)$ . Como se ha mencionado en el capítulo 2, Newell defendió el concepto de inteligencia porque una teoría de los comportamientos tiene que tener una noción de una habilidad o potencia que muestre una escala. En el caso de la fasa, cada valor de fasa implica la aparición de diferentes exocomportamientos; pero hablar de una escala solo es posible cuando se restringe el conjunto de exocomportamientos a los de un entorno, ya que el conjunto de exocomportamientos no tienen una relación de orden estricto que los relacione. Esto es así, porque existen diversidad de entornos en la naturaleza y la adecuación biológica a cada uno implica diferentes comportamientos.

#### 4.4.0.1. La interpretación de la fasa

Siguiendo las reglas de razonamiento establecidas en la sección 4.2.2 se ha concluido que los sistemas exoactivos poseen una propiedad que causa los exocomportamientos. Sin embargo, las reglas no explican cómo interpretar la propiedad, cómo describirla ni cómo describir la relación entre el fenómeno y la causa. Ante esta situación, es interesante considerar las investigaciones realizadas hasta el día de hoy sobre los comportamientos. El principio fundamental de la ciencia cognitiva es que los procesos cognitivos son procesos computacionales de procesamiento de información [430, 381]. También

sabemos que comportamientos en los que no hay representaciones mentales pueden ser producidos por procesos computacionales [85, 86, 83]. Además, aunque la máquina de Turing es un modelo computacional determinista, a partir de los años 60 se empezaron a investigar los procesos computacionales probabilísticos. En 1963, Michael O. Rabin dio la definición de automata probabilístico [475]. A eso hay que añadir que desde los trabajos Eugene S. Santos dando una definición de máquina de Turing probabilista [502]. Así, actualmente se entiende que los procesos probabilistas también son procesos computacionales. Este hecho es importante porque, aunque en los comportamientos aleatorios de microorganismos o animales no haya un procesamiento de información, sí corresponden con procesos computacionales. Uniendo estos dos hechos, se puede concluir que la fasa es un proceso computacional.

Ahora bien, puesto que el *explanans* que se busca debe cumplir con la hipótesis del fisicalismo de los comportamientos, la fasa de un sistema exoactivo no puede tratarse de una entidad matemática del universo de las ideas de Platón sino debe ser una entidad física. Dentro de las opciones fisicalistas, no puede tratarse de una propiedad de la materia porque solo se encuentra en determinados sistemas físicos. Además, no necesariamente es una propiedad de todo el sistema porque en los organismos multicelulares la desaparición de una parte del sistema no implica que el sistema deje de ser un sistema exoactivo ni que su valor de fasa cambie. De todo lo anterior y puesto que un sistema físico está compuesto de subsistemas se ha de concluir que la fasa de un sistema exoactivo se corresponde con un subsistema físico propio o impropio de él. Ese subsistema lleva a cabo el proceso de computación que genera el exocomportamiento del sistema exoactivo, aunque este subsistema no tiene que ser un sistema con frontera. Al subsistema que se corresponde con la fasa de un sistema exoactivo se le denotará por  $\mathcal{F}$ . Ese subsistema causa los cambios en las cantidades de energía que conforman el exocomportamiento. Es decir, la secuencia de un exocomportamiento,  $\varepsilon$ , es determinada por una secuencia de señales producidas por  $\mathcal{F}$ . Se denotará por  $\mathbb{F}$  al conjunto de todos los valores de fasa de la naturaleza, siendo  $\mathcal{F}$  un elemento de él.

Hay que notar que  $\mathcal{F}$  es todo lo que controle los circuitos y fuentes de energía del sistema exoactivo. Por ejemplo, en los animales  $\mathcal{F}$  no es solo las neuronas del sistema nervioso, además los receptores de acetilcolina de los músculos forman parte de  $\mathcal{F}$ , ya que el control del uso de energía en los músculos es determinado por esos receptores.

En la definición de los exocomportamientos se había visto que los procesos ontogénicos son exocomportamientos. Una de las conclusiones del capítulo 3 fue que el desarrollo de un fenotipo consiste en un proceso llevado a cabo por mecanismos genéticos y moleculares que transforman parte de la información almacenada en el genotipo en el fenotipo. En base a ese hecho, se puede considerar que un proceso ontogénico es un proceso computacional. Esta afirmación se puede apoyar con otros dos argumentos. El primer argumento es que en el juego de la vida de John Conway [172] en el que se puede observar un equivalente a procesos ontogénicos, los patrones matusalenes que cambian durante una gran número de turnos. El juego de la vida es un autómata

celular, y por lo tanto se trata de procesos computacionales. El segundo argumento es que los mecanismos genéticos y moleculares que tienen los organismos biológicos se están usando para crear mecanismos de computación [96, 63]. Pese a que observando el comportamiento de un mamífero y su proceso ontogénico se podría concluir que se trataran de fenómenos diferentes, si observa el proceso ontogénico de un organismo unicelular que consiste en su capacidad de cambiar de fenotipo la distinción entre su comportamiento y su proceso ontogénico desaparece. Así, en base a la interpretación de fasa, y considerando que los procesos ontogénicos son procesos computacionales, se puede concluir que el conjunto de los exocomportamientos contiene a los procesos ontogénicos es un conjunto de fenómenos que existe en la naturaleza.

#### 4.4.1. Análisis matemático de la fasa

Puesto que se ha propuesto la existencia de un conjunto de valores  $\mathbb{F}$  que causan los exocomportamientos, el siguiente paso ha sido investigar que subconjuntos existen dentro de  $\mathbb{F}$  para entender mejor en que consiste la fasa. Partiendo del hecho de que la fasa se corresponde con un sistema computacional, se puede conocer más sobre los subconjuntos de valores de fasa a partir de su análisis matemático. La teoría algorítmica de la información busca encontrar una descripción,  $\sigma$ , de una cadena,  $s$ . En la teoría algorítmica de la información una función parcial computable que asocia cadenas de símbolos a cadenas de símbolos es denominada *máquina*. Si  $M$  es una máquina y se cumple que  $M(\sigma) = s$  entonces se dice que  $\sigma$  es una descripción de  $s$ , o un programa para generar  $s$ . Un exocompartamiento es una secuencia de cantidades de energía fluyendo por los circuitos del sistema que actúan sobre propiedades observables y medibles del sistema exoactivo. Cada una de las cantidades de energía puede ser descrita como un símbolo. En este análisis, el intervalo de tiempo se ignorará ya que la duración de los intervalos no afecta a la naturaleza computacional de la secuencia.

En la teoría algorítmica de la información se considera que la máquina es una máquina universal de Turing, ya que lo que interesa es estudiar las descripciones  $\sigma$  para generar la secuencia  $s$ . Sin embargo, lo que interesa para analizar la fasa es conocer tanto las posibles máquinas  $M$ , como las posibles descripciones  $\sigma$ . Esto es así porque cada  $M$  y  $\sigma$  constituyen un valor de fasa ya que generarán una secuencia, y esa secuencia describe un exocomportamiento  $\varepsilon$ . Por lo tanto, el valor de fasa de un sistema exoactivo queda descrito por la máquina  $M$  y la entrada  $\sigma$  que generen la secuencia de símbolos isomorfa al exocomportamiento del sistema. Así, en la TGE, ni  $M$  es una máquina universal de Turing, ni  $\sigma$  es una codificación de una máquina de Turing. Así, partiendo de unas nociones generales de  $M$  y  $\sigma$  el objetivo es analizar estos dos elementos matemáticamente en un contexto computacional para obtener una descripción de la fasa.

Consideremos, las siguientes generalizaciones:

- $\sigma = \langle p, d, i \rangle$  donde  $p$  es una cadena que codifica una máquina de Turing,  $d$  una cadena que codifica un conjunto de datos, e  $i$  es una cadena que contiene las interacciones que recibirá del entorno la máquina.
- $M$  es una máquina cualquiera de computación; puede ser determinista o probabilista; puede ser de computación cerrada o iterativa; puede ser universal o específica.

Partiendo de las dos generalizaciones anteriores se analiza las distintas posibilidades que surgen para generar un exocomportamiento. Considerando el caso en el que  $\sigma$  es una cadena vacía, que denotamos por  $|\sigma| = 0$ . Es decir,  $M$  genera exocomportamientos sin que exista ningún tipo de memoria que contenga distintos valores o instrucciones  $M = \varepsilon$ . En este caso existen dos posibilidades. Una opción es que  $M$  es determinista. La otra opción es que  $M$  es probabilista. Cada uno de estos dos casos muestra un tipo de fasa diferente, ya que en cada uno de los casos lo único que determina  $\varepsilon$  es la naturaleza de la máquina  $M$ . En el caso de que  $M$  sea probabilista, la secuencia del exocomportamiento es una secuencia aleatoria. En base a ello, se determina que existe un tipo de fasa que se denominará *aleatoria*, y el conjunto de valores de fasa de este tipo se denotará por  $\mathbb{F}_\alpha$ . En el otro caso, siendo  $M$  determinista, cada elemento de la secuencia del exocomportamiento es función de términos anteriores. Así, a este tipo de fasa se la denominará *recurrente*, y el conjunto de valores de fasa de tipo recurrente se denotará por  $\mathbb{F}_\beta$ .

Ahora, consideremos las opciones posibles siendo  $\sigma > 0$ :

- Siendo  $|p| > 0$ ,  $|d| = 0$ , y  $|i| = 0$  entonces  $M$  se trata de una máquina universal. Aunque la máquina universal permite que se puedan generar diferentes  $\sigma$  con diferentes  $p$ , para cada programa  $p$  existe una máquina (que se trata de la máquina codificada por  $p$ ) que por ella misma genera  $\sigma$ . Por lo tanto, se cumple que

$$M(p) = M_p = \varepsilon$$

así que  $M(p)$  no genera un  $\varepsilon$  de naturaleza diferente del sistema exoactivo con fasa del tipo recurrente o aleatorio. En este caso,  $M$  es una máquina universal, pero su tipo de fasa será recurrente o aleatorio dependiendo del tipo de  $M$ , y la naturaleza de  $M(p)$  es la misma que la de  $M_p$ .

- Siendo  $|p| = 0$ ,  $|d| > 1$ , y  $|i| = 0$  se trata de una generalización que permite a una misma máquina  $M$  generar distintas secuencias. Igual que en el caso anterior, para cada una de esas secuencias existe una máquina  $M_d$  que puede generarla también, ya que  $M(d)$  tiene una variable mientras que  $M_d$  tiene una constante. Por lo tanto, tener los datos cableados en el hardware o disponer de una memoria para almacenarlos no hace que la naturaleza de la fasa cambie.

- Siendo  $|p| = 0$ ,  $|d| = 0$ ,  $|i| = |\varepsilon|$ , y que por cada  $i_x$  perteneciente a  $i$  en el exocomportamiento  $\varepsilon$  siempre se produce en la secuencia el mismo  $c_x$ , es decir,

$$M(\dots, i_x, \dots, i_y, \dots) = \dots, c_x, \dots, c_y \text{ tal que si } i_x = i_y \Rightarrow c_x = c_y$$

entonces  $\varepsilon$  es directamente producido por las interacciones que recibe por parte del entorno. Nótese que, por las condiciones impuestas,  $M$  es determinista ya que si hubiera aleatoriedad no se cumpliría que para cada  $i_x$  siempre se genere el mismo  $c_x$ . La naturaleza de este tipo de  $\varepsilon$  es diferente de la que se pueden generar por fasa recurrente o aleatoria. Por lo tanto, se concluye que hay otro tipo de fasa a la que se denominará reactiva, y el conjunto de valores de fasa de tipo reactivo se denotará por  $\mathbb{F}_\gamma$ .

Así, se tiene que:

$$\mathbb{F}_\alpha, \mathbb{F}_\beta, \mathbb{F}_\gamma \subset \mathbb{F}$$

Una cuestión que puede surgir es si es compatible la fasa reactiva con que un sistema sea exoactivo, ya que un sistema exoactivo es no endodirigido ni endoregulado. La respuesta a esa cuestión es que no hay contradicción, ya que en la fasa reactiva la energía de las señales que recibe no es capaz de modificar por sí misma los flujos de energía del sistema exoactivo. El sistema exoactivo tiene que amplificar la señal de entrada para que provoque cambios en los flujos de energía del sistema. Por lo tanto, un sistema físico puede cumplir el no ser endodirigido ni endoregulado, y que su fasa sea de tipo reactivo.

Otra cuestión que emerge sobre el resultado del análisis de que existen diferentes tipos de fasa fundamental es qué tipo de relación existe entre ellos de cara a producir un exocomportamiento. Por lo que se sabe de otras propiedades físicas de la materia, hay diferentes posibilidades. Por ejemplo, una opción es que se pueden combinar como ocurre en la carga de color de los quarks. Otra opción es que se anulen entre ellos como ocurre en la carga eléctrica. Si tomamos en cuenta la formulación general que se ha hecho para realizar el análisis tenemos que concluir que los tipos de fasa no se anulan o se invalidan entre ellos. El análisis se ha centrado en aislar características que puede tener la fasa de un sistema exoactivo anulando el resto. Pero si no se anulan, las características pueden darse todas a la vez. Por lo tanto, el tipo de un valor de fasa puede ser una combinación de los tipos fundamentales. Así, en el conjunto se cumplirá que :

$$\mathbb{F}_{\alpha\beta}, \mathbb{F}_{\beta\gamma}, \mathbb{F}_{\alpha\beta\gamma} \subset \mathbb{F}$$

El siguiente postulado contiene los resultados del análisis matemático que se ha hecho sobre el concepto de fasa.

**Postulado 2. (*Principio del carácter de los valores de fasa*):** *El valor de fasa de un sistema exoactivo pertenece, o a uno de los tres tipos fundamentales (aleatorio, secuencial o reactivo), o a una combinación de ellos.*

El principio quedaría formulado de la siguiente manera:

$$\mathcal{F} \in \mathbb{F} = \mathbb{F}_\alpha \cup \mathbb{F}_\beta \cup \mathbb{F}_\gamma \cup \mathbb{F}_{\alpha\beta} \cup \mathbb{F}_{\beta\gamma} \cup \mathbb{F}_{\alpha\beta\gamma}$$

El método científico determina que una teoría científica necesita que sus afirmaciones sean falseables. Esto implica que para cada afirmación que hace la teoría deben existir experimentos reproducibles en los que entre sus posibles resultados exista alguno que de obtenerse implicaría que la afirmación es falsa. El principio del carácter de los valores de la fasa es falseable, ya que si se encuentra un modo físico para generar una secuencia con un método matemático diferente de aquellos establecidos por el postulado el principio se demostraría falso.

Debe notarse que al hacerse el análisis matemático de la fasa se ha considerado  $M$  desde un punto de vista funcional. Sin embargo, un exocomportamiento no es solo una sucesión de estados sino también está conformado por la duración de esos estados. Esto implica que un valor de fasa no queda completamente descrito mediante una función computacional, ya que las características físicas de  $M$  afectan a la duración de los intervalos de tiempo de niveles de energía del exocomportamiento. Por ejemplo, un sistema exoactivo cuyas características le permitan computar más rápido generará exocomportamientos diferentes de otros sistemas que tiene propiedades físicas que le hacen computar más despacio. Por lo tanto, un valor de fasa  $\mathcal{F}$  es determinado por: un modelo computacional, los datos que usa el sistema computacional y las características físicas de la implementación del modelo computacional.

## 4.5. Identificación experimental de tipos de fasa fundamental

La TGE dicta que el exocomportamiento de un sistema exoactivo es producido por su valor de fasa. La TGE también dicta que existen tres tipos fundamentales de fasa. Puesto que la TGE es una teoría científica debe ser posible identificar experimentalmente que un valor de fasa pertenece a un tipo fundamental y no a otro. Sin embargo, un exocomportamiento es una secuencia finita, de manera que no se puede decidir qué tipo de fasa la ha generado observándola únicamente, ya que una misma secuencia puede ser producida por distintos métodos computacionales. Pero teniendo en cuenta que cada tipo de fasa fundamental tiene características que la distinguen de las demás, esas características deben ser observables en los exocomportamientos que genere una población de sistemas exoactivos iguales. Nótese que al ser iguales los

sistemas exoactivos tienen que tener el mismo valor de fasa. Por lo tanto, se puede conocer el tipo de fasa que tiene un sistema exoactivo colocando una población de sistemas exoactivos iguales al que se desea estudiar su tipo de fasa en entornos diferentes. Analizando los exocomportamientos de los individuos de las poblaciones se obtendrán las características del tipo de fasa que contiene el sistema exoactivo. A continuación, se deducen criterios experimentales para determinar los tipos de fasa fundamentales a partir de una población de sistemas exoactivos iguales.

- Criterio para la fasa aleatoria. El análisis matemático mostró que existe un tipo de fasa fundamental cuya naturaleza es aleatoria que consiste en un sistema computacional cerrado que genera una secuencia entre un conjunto de ellas basado únicamente en aleatoriedad. Por lo tanto, si se coloca una población de sistemas exoactivos iguales en un entorno isotrópico, los exocomportamientos tienen que ser diferentes si el valor de su fasa es de tipo aleatorio. Esto se deduce de los dos siguientes descartes. Se descarta la fasa recurrente porque al ser todos los sistemas exoactivos iguales su valor de fasa es el mismo, pero al ser exocomportamientos diferentes no se está generando la misma secuencia. También se descarta la fasa reactiva porque al ser un entorno isotrópico si fuera fasa reactiva todos los sistemas de la población tendrían que reaccionar igual pero no lo hacen.
- Criterio experimental para la fasa recurrente. En el análisis matemático que se ha realizado se ha hallado que existe un tipo de fasa fundamental de naturaleza recurrente que consiste en un sistema computacional cerrado que genera una secuencia a través de un proceso recurrente. Por lo tanto, un sistema exoactivo con un tipo de fasa recurrente debe generar una secuencia específica independientemente del tipo de entorno en el que se encuentre. Por lo tanto, si se coloca una población de sistemas exoactivos iguales en un entorno anisotrópico, lo que implica que cada uno se encuentra sumergido en condiciones de entorno distintas; pero todos generan la misma secuencia, se está ante un valor de fasa de tipo recurrente. Esto se deduce porque por un lado se descarta el tipo de fasa reactiva porque no está reaccionando a las diferencias del entorno, y por otro lado, la secuencia es siempre la misma en todos los individuos por lo que tampoco puede tratarse de fasa aleatoria, ya que deberían aparecer variaciones en los exocomportamientos y no lo hace.
- Criterio para la fasa reactiva. El análisis que se ha realizado ha mostrado que existe un tipo de fasa fundamental denominada reactiva que consiste en un sistema computacional abierto que genera una secuencia que es función de los valores de ciertas características del entorno. A diferencia de los dos casos anteriores no se puede deducir el tipo de fasa introduciendo la población de sistemas exoactivos en un solo entorno, ya que en un único entorno anisotrópico no se podría distinguir entre fasa aleatoria y fasa reactiva, y si fuera isotrópico no se podría

distinguir fasa reactiva y secuencial. El criterio para distinguir la fasa reactiva es dividir la población y colocar cada grupo en un entorno isotrópico diferente. Si en cada entorno, cada uno de los grupos genera el mismo exocomportamiento, pero los exocomportamientos de sistemas de grupos distintos son diferentes, entonces su fasa es de tipo reactiva. Esto se deduce de los dos descartes siguientes. Se descarta la fasa recurrente porque al ser todos los sistemas exoactivos iguales el exocomportamiento no depende del entorno por lo que en los dos grupos debería ser igual. También se descarta la fasa aleatoria porque al ser cada uno de los entornos isotrópico no hay nada que pudiera provocar que hubiera diferencias, por lo que no pueden aparecer exocomportamientos diferentes dentro del mismo grupo.

En el análisis matemático del apartado anterior se obtuvo que existen tres tipos de fundamentales de fasa. Ahora se han obtenido criterios experimentales para determinar cada uno de los tres tipos fundamentales de fasa hallados en el análisis matemático. En base a ellos es posible demostrar experimentalmente que los tres tipos fundamentales de fasa existen y que el conjunto de valores de cada uno de ellos no es un conjunto vacío. Esto queda afirmado en el siguiente teorema:

**Teorema 1** (Teorema fundamental de la fasa). *Ningún conjunto de valores de fasa fundamental es vacío.*

La demostración de este teorema consiste en probar que cada uno de los conjuntos es no vacío. Es decir, hay que mostrar que al menos un valor de fasa pertenece a cada tipo fundamental. Esto se puede lograr haciendo que un sistema exoactivo genere un exocomportamiento con un valor de fasa de cada tipo fundamental. La manera más sencilla de hacer esto es con robots, ya que se dan dos circunstancias. Una es que los robots son una herramienta para estudiar el comportamiento [345]. La otra es que, ya que la TGE es una teoría que abarca tanto sistemas biológicos como los no biológicos, los robots son sujetos del dominio de la TGE. Por lo tanto, los robots son elementos para la experimentación y contrastación directa de la TGE. De esta manera, probar el teorema consiste en probar tres lemas, donde cada lema afirma que existe al menos un elemento en cada uno de los conjuntos de valores de fasa fundamental. Para probar cada uno de esos lemas, se da un programa que generará un exocomportamiento que pertenece a cada uno de los tipos fundamentales de fasa. Recuérdese que el valor de fasa se corresponde con el modelo de computación, los datos que contiene y sus características físicas, por lo que la arquitectura computacional del robot y el programa que cargue conforman el valor de fasa que posee. Así, creando un programa que sea cargado y ejecutado por una población de robots al encenderse, los robots están produciendo exocomportamientos. El tipo de fasa puede ser identificado mediante el análisis de los exocomportamientos mediante los criterios de identificación experimentales de fasa.

**Lema 1.** *El conjunto de valores de la fasa aleatoria ( $\mathbb{F}_\alpha$ ) es no vacío.*

*Demostración.* Supóngase que se tienen 10 robots que pueden realizar cuatro acciones: avanzar, retroceder, girar a la izquierda 5 grados y girar a la derecha 5 grados. Cada una de esas acciones es denotada por 1,2,3,4. El siguiente programa será almacenado en la memoria de los 10 robots.

```

program aleatorio;
var accion:integer; // variable para almacenar el identificador de una acción
begin
while true do // un bucle que se va a ejecutar infinitamente
begin
accion:= random(4); // elige un valor aleatoriamente entre 1 y 4 ambos incluidos
execute(accion);
end;
end.

```

Esos diez robots son colocados en una misma habitación isotrópica. Tras encenderse los diez robots, cargan y empiezan la ejecución del programa. La ejecución del programa en cada robot provoca que cada uno de ellos ejecute una secuencia de acciones diferentes. Eso muestra que los robots tienen un valor de fase aleatoria. Por lo tanto, se demuestra que el conjunto de valores de la fase aleatoria no es un conjunto vacío porque se ha mostrado que al menos tiene un elemento. La clave de esta prueba es la función random. Si el robot dispone de un verdadero generador de números aleatorios [147, 199], la población mostrará obligatoriamente exocomportamientos que cumplen los criterios para identificar fase aleatoria. Si los robots solo disponen de un generador de números pseudoaleatorios el hecho importante es que la semilla de cada generador de números aleatorios sea diferente y los exocomportamientos cumplirán también los criterios para identificar fase aleatoria.  $\square$

**Lema 2.** *El conjunto de valores de la fase recurrente,  $\mathbb{F}_\beta$ , es no vacío.*

*Demostración.* Supóngase que se tienen los 10 robots anteriores. En este caso, el siguiente programa que será almacenado en la memoria de los 10 robots es el siguiente:

```

program recurrente;
var a,i:integer;
i:=1;
while true do
begin

case i of
1: a:= 1;
2: a:= 1;
3: a:= 1;

```

```

4: a:= 3;
5: a:= 1;
6: a:= 3;
7: a:= 2;
8: a:= 2;
9: a:= 4;
end;
execute(a);
i:= i+1;
if i>9 then i:=1;
end;

```

Los diez robots son colocados en una misma habitación anisotrópica. Tras encenderse los diez robots, cargan y empiezan la ejecución del programa. La ejecución del programa provoca que todos los robots ejecuten la misma secuencia de acciones independientemente de las coordenadas de la posición de origen de cada robot. Esto cumple el criterio para identificar fase recurrente. Por lo tanto, se demuestra que el conjunto de valores de la fase recurrente es un conjunto no vacío se ha encontrado un valor que pertenece a él.  $\square$

**Lema 3.** *El conjunto de valores de la fase reactiva  $\mathbb{F}_\gamma$  es no vacío.*

*Demostración.* Supongase que se tienen los 10 robots de los casos anteriores de nuevo. El siguiente programa será almacenado en la memoria de los 10 robots.

```

program reactivo
var t: integer; // almacena la temperatura
    a:integer;

while true do
begin
t:= register_temperature; // registra la temperatura del exterior
if t > 25
then
    a:= 3 // escoge girar a la izquierda
else
    a= 4; // escoge girar a la derecha
execute(a);
end;

```

Los diez robots se dividen en dos grupos de 5. Cada grupo es colocado en una habitación isotrópica, pero en una la temperatura es mayor de 25 grados y en otra es menor de 25 grados. Tras encenderse los diez robots, cargan y empiezan la ejecución

del programa. La ejecución del programa por parte de los robots provoca que los robots de cada habitación ejecuten la misma secuencia de acciones, pero la secuencia de los de una habitación es diferente de la secuencia de los de la otra, ya que una gira en un sentido, y los de la otra giran en sentido contrario. Los exocomportamientos cumplen los criterios que tienen un valor de fase reactiva. Por lo tanto, el conjunto de fase reactiva es un conjunto no vacío porque se ha mostrado que al menos tiene un elemento.  $\square$

Puesto que se han probado los tres lemas, el teorema fundamental de la fase también queda probado, ya que ninguno de los tipos fundamentales de fase es no vacío.

## 4.6. Leyes de la fase y predicciones para sistemas exoactivos

Una teoría científica debe dar explicaciones y hacer predicciones de fenómenos. Para ello, una teoría debe tener leyes que relacionen la causa con fenómenos para que sea capaz de dar explicaciones y realizar predicciones. En el capítulo 3 se vio que el comportamiento afecta al éxito reproductivo determinando la supervivencia o el emparejamiento. Así, el comportamiento es un rasgo del fenotipo que interviene en la supervivencia, o la extinción, de una especie por selección natural. Usualmente, se ha buscado que las teorías sobre algún conjunto de comportamientos sean capaces de predecir algún comportamiento de los individuos. La TGE haría las mismas predicciones que la psicología cognitiva o el conductismo, ya que como se ha dicho la TGE interpreta cada uno de los mecanismos de esas propuestas como un valor de fase. La cuestión es si la TGE puede hacer alguna predicción o producir alguna explicación no dada anteriormente. Considerando el hecho mencionado de la importancia del comportamiento, la capacidad explicativa y predictiva de la TGE debe verse en fenómenos de supervivencia o extinción de poblaciones en base al exocomportamiento de sus individuos. Dado que en la teoría de la evolución darwinista no existe un rasgo, o un conjunto de rasgos que conviertan a un individuo en el mejor adaptado de manera universal, la TGE debe dar explicación y hacer predicciones sobre que valores de fase poseen mayor adecuación biológica a los distintos tipos de entorno. Puesto que la TGE abarca comportamientos que se pensaban que no eran seleccionados por la naturaleza, la selección de esos comportamientos es un fenómeno que debe ser explicado. Así, la TGE formula leyes para explicar la adecuación biológica de los organismos en los entornos en los que existe un proceso de selección natural. Dado que cada entorno natural determina diferentes valores de adecuación biológica, las características de un entorno determinarán los valores de adecuación de los sistemas que son sometidos al proceso de selección. Se denotará al valor de adecuación biológica de una unidad de selección con el símbolo  $W$ . El estado de un entorno natural cambia con el tiempo. En matemáticas los objetos cuyo estado cambia con el tiempo son considerados sistemas dinámicos. Así, un entorno natural

desde el punto de vista matemático es un sistema dinámico. La secuencia de estados de un sistema dinámico se llama órbita. Usualmente un entorno natural posee distintos nichos. Entendiéndose nicho como el conjunto de características del entorno que son relevantes para la vida de los organismos. Así, un elemento importante para relacionar procesos evolutivos con la TGE será la órbita del nicho de los sistemas exoactivos. Cuando el tiempo que tarda en cambiar el nicho de un estado a otro es menor que el tiempo que le llevaría al sistema detectar el cambio y dar una respuesta se dirá que la órbita del nicho es *abrupta*; pero si el tiempo que tarda en suceder el cambio es mayor que el que necesita el organismo para detectarlo y reaccionar se dirá que su órbita es suave. Nótese que en un mismo entorno puede haber nichos con orbitas abruptas y orbitas suaves.

A continuación, se van a presentar las leyes que se proponen en la formulación actual de la TGE para relacionar la fasa con fenómenos evolutivos.

**Ley 1** (Ley de la fasa aleatoria). *Sea un entorno con selección natural,  $E$ , en el que existen un conjunto de poblaciones de sistemas exoactivos que solo difieren en el valor de fasa que posee cada población  $P_{\mathcal{F}}$ . Si la órbita del nicho es no cerrada y abrupta o una secuencia aleatoria de estados, el valor de fasa que logrará que una población tenga el valor de adecuación más alto pertenecerá al tipo de fasa aleatoria.*

$$\arg \max_{\mathcal{F} \in \mathbb{F}} W(P_{\mathcal{F}}, E) \in \mathbb{F}_{\alpha}$$

La aleatoriedad en la aparición de estados dañinos del entorno para los sistemas exoactivos impide que cada individuo pueda predecir la aparición de los estados. La ley de la fasa aleatoria explica los comportamientos observados en microorganismos cuyas decisiones aleatorias les permite ser los mejor adaptados a su nicho [40, 281]

**Ley 2** (Ley de la fasa recurrente). *Sea un entorno con selección natural,  $E$ , en el que existen un conjunto de poblaciones de sistemas exoactivos que solo difieren en el valor de fasa que posee cada población  $P_{\mathcal{F}}$ . Si la órbita del nicho es cerrada y suave, el valor de fasa que logra que una población tenga el valor de adecuación más alto pertenece al tipo de fasa recurrente.*

$$\arg \max_{\mathcal{F} \in \mathbb{F}} W(P_{\mathcal{F}}, E) \in \mathbb{F}_{\beta}$$

La ley de la fase recurrente, además también de hacer predicciones, explica los descubrimientos sobre ritmos circadianos [193, 358].

**Ley 3** (Ley de la fasa reactiva). *Sea un entorno con selección natural en el que existen un conjunto de poblaciones de sistemas exoactivos que solo difieren en el valor de fasa que posee cada población  $P_{\mathcal{F}}$ . Si la órbita del nicho es suave y no es cerrada, el valor de fasa que logra que una población tenga el valor de adecuación más alto pertenece al tipo de fasa reactiva.*

$$\arg \max_{\mathcal{F} \in \mathbb{F}} W(P_{\mathcal{F}}, E) \in \mathbb{F}_{\gamma}$$

La ley de la fasa reactiva, además de permitir hacer predicciones, explica que microorganismos desarrollen complejos mecanismos de procesamiento de la información [239][396].

Las tres leyes enunciadas son afirmaciones falseables, ya que, al predecir un mayor valor de adecuación biológica, tendrían que darse dos fenómenos. Uno la esperanza media de la población con el valor de fasa que predice la ley tiene que ser mayor que las poblaciones con otros. Segundo, si los sistemas exoactivos son unidades evolutivas, la diferencia reproductiva debería conducir a una evolución que extinga a los que la ley no les asigna el mayor valor de adecuación y sobrevivan los que ley predice. El mayor problema para realizar experimentos que pongan a prueba las leyes es conocer el valor de fasa de los individuos de las poblaciones de sistemas exoactivos, pero dado que la TGE abarca sistemas no biológicos, los experimentos pueden realizarse mediante robots que permitan conocer perfectamente el valor de los sistemas del experimento.

## 4.7. Reinterpretación de investigaciones previas sobre comportamientos

Una cuestión que merece ser examinada es qué papel juegan las investigaciones y resultados obtenidos hasta el momento sobre los comportamientos si se acepta la TGE. Lo primero que se debe de mencionar es que de la misma manera que la mecánica de Newton no invalidó los trabajos de Galileo ni los de Kepler, sino que permitió que se tuviera una nueva visión sobre ellos, la TGE es una teoría que aporta una nueva visión sobre las investigaciones y resultados que se han hecho hasta el día de hoy sobre los comportamientos. Por ejemplo, cognitivismo y conductismo han sido dos teorías enfrentadas por explicar los comportamientos de los seres humanos. Esa batalla la ganó el cognitivismo mostrando que el conductismo no puede explicar determinados comportamientos humanos [430]. Sin embargo, el conductismo en IA ha mostrado su efectividad para lograr generar comportamientos [86, 581, 83]. Mientras conductismo y cognitivismo han sido consideradas dos teorías opuestas para explicar los comportamientos, dentro del marco de la TGE son dos conjuntos de valores de fasa:  $\mathbb{F}_{cond}$  y  $\mathbb{F}_{cog}$ . Cumpliéndose que:

$$\mathbb{F}_{cond}, \mathbb{F}_{cog} \subset \mathbb{F}_\gamma$$

Esta visión permite entender que el conductismo haya permitido que la IA obtenga determinados éxitos en robótica. Desde el punto de vista de la TGE, para ciertos entornos un valor de fasa del conjunto de los conductistas puede dar el mayor valor de adaptabilidad a ese entorno. Además, la TGE también acoge la propuesta de que una combinación de ambos mecanismos puede generar el comportamiento [570], ya que los valores de fasa siempre pueden ser combinados. Dándose que:

$$\mathbb{F}_{cond} \cup \mathbb{F}_{cog} \neq \mathbb{F}_\gamma$$

Respecto a las arquitecturas cognitivas, la TGE las interpreta como valores de fasa. El programa de teorías unificadas de Newell supone que hay solo una manera en la que el comportamiento inteligente puede generarse. Sin embargo, existe una diversidad de arquitecturas cognitivas que ha estado creciendo. Ese hecho contradice el principio en el que se basa el programa de investigación de Newell que se irían eliminando entre ellas. En cambio, ese mismo hecho puede ser perfectamente asimilado en la TGE, ya que la existencia de múltiples arquitecturas cognitivas es compatible con la relación entre exocomportamiento y fasa de la TGE. Esto es debido a que la TGE no dice cuántos valores de fasa son solución para generar un exocomportamiento. Es decir, usando la ecuación 4.1, la existencia de múltiples arquitecturas cognitivas para generar los comportamientos cognitivos se describiría en la siguiente desigualdad:

$$\{\mathcal{F} : \varepsilon = C(\mathcal{F})\} \geq 1$$

Los resultados de investigación en cognición comparativa han llevado a proponer que no se debe considerar la existencia de una única forma auténtica de cognición, sino que existe una diversidad de ellas, y que todas debería ser consideradas en un único marco que las contenga [597]. TGE puede ser perfectamente ese marco, ya que cada forma de cognición puede ser interpretada como un valor o conjunto de valores de fasa.

Respecto a las diferencias individuales y las teorías sobre la inteligencia, tales como la hipótesis de la mielinización, la hipótesis de la poda neuronal o la hipótesis de la plasticidad, la TGE interpreta las variables físicas que mencionan esas hipótesis como variables que determinan un conjunto de valores de fasa que son ligeramente diferentes unos de otros. Esa variación en los valores de la fasa de los sistemas causaría diferencias en los exocomportamientos, y así se apreciarían las diferencias individuales en los organismos.

## 4.8. Análisis de la investigación sobre la TGE

Desde los comienzos del siglo XX la inteligencia ha sido el concepto entorno al que han girado los intentos por explicar los comportamientos complejos de los animales. Sin embargo, no ha sido posible formular una teoría sólida de qué es la inteligencia. Además, las definiciones dadas hasta ahora hacen que sea imposible que pueda ser usado como elemento que permita una unificación de los comportamientos, ya se plantea en ellas como un elemento diferenciador. Ante esta situación, se ha llevado a cabo una investigación para formular una teoría que unifique los comportamientos. En este capítulo, ha presentado la esa investigación y su resultado que es una teoría denominada teoría

general del exocomportamiento. La TGE formula un único conjunto de fenómenos, denominados exocomportamientos, que incluye como fenómenos los comportamientos de sistemas biológicos, no biológicos autónomos y los procesos ontogénicos. Hasta ahora la unificación de comportamientos de sistemas biológicos y máquinas ha sido abordada en diversas ocasiones. Sin embargo, los procesos ontogénicos nunca se han considerado por parte de la psicología o la IA. Tan sólo desde el punto de vista evolutivo se ha especulado sobre la posible conexión entre la plasticidad fenotípica y los comportamientos de seres vivos. Por lo tanto, la TGE propone una visión completamente nueva de cómo considerar esos tres tipos de fenómenos.

La unificación no es una cuestión metafísica o filosófica. Cada teoría de unificación supone un importante paso en la ciencia porque proporciona una nueva y más completa forma de comprender la naturaleza. Teorías de unificación como la teoría celular que unificó el reino animal y vegetal; la teoría atómica moderna que unificó la química orgánica e inorgánica; la teoría electromagnética que unificó la electricidad y el magnetismo; la mecánica de Newton que unificó los movimientos celestes y los terrestres; y la relatividad general que unificó la masa inercial y la gravitatoria; son ejemplos de las consecuencias que tiene para la ciencia el que se obtenga una teoría de unificación. La cuestión de la unificación es una cuestión presente también en las ciencias cognitivas [13]; aunque la TGE propone una unificación más general de lo que por intentos anteriores se podía prever. Una unificación se basa en encontrar que en los fenómenos interviene una propiedad en común, aunque debido a las diferencias de las situaciones los efectos son tan diferentes que parecen ser fenómenos distintos. Por ejemplo, la carga eléctrica en reposo y en movimiento provocan fenómenos de características diferentes. En el caso de la TGE, la definición de exocomportamiento es la clave de la unificación, ya que pone de manifiesto que los comportamientos de sistemas biológicos, de sistemas no biológicos autónomos y los fenómenos de plasticidad fenotípica tienen en común que suceden porque el sistema dirige y controla sus flujos energía. Así, puesto que todos son los mismos fenómenos tienen que tener la misma causa. De esa conclusión, nace la TGE.

En la investigación sobre la TGE se han abordado diversas cuestiones sobre los elementos que la componen y cómo se relacionan con los exocomportamientos. Uno de los resultados inesperados de la formulación de la TGE es que los procesos ontogénicos de los organismos biológicos pertenecen al mismo conjunto que los comportamientos de sistemas biológicos y no biológicos autónomos. Este hecho es inesperado porque por ejemplo aparentemente el comportamiento de un mamífero parece un fenómeno muy diferente de su proceso ontogénico. El análisis ha mostrado que la causa de uno y otro proceso es la misma; pero donde puede observarse esa conclusión que emerge de la TGE es en los organismos unicelulares. En los organismos unicelulares, la línea que separa su comportamiento y su proceso ontogénico desaparece. En los años 90 dos investigaciones independientes encontraron un mismo modelo matemático evolutivo. Uno era un modelo que explicaba las ventajas del control plástico del desarrollo [401].

El otro era sobre las ventajas del aprendizaje [538]. La TGE afirma que los puntos en común entre comportamientos y procesos ontogénicos no son coincidencias, sino que hay una conexión profunda entre ellos. Otra cuestión que se ha abordado es cuantos tipos de fasa existen. Se ha visto mediante un análisis matemático que existen tres tipos de fasa fundamentales y que estos tipos se pueden combinar entre ellos. Un hecho importante es que, tanto en procesos ontogénicos como en los comportamientos de organismos, si son observados como exocomportamientos se ve que les hay de los tres tipos de fasa fundamental. Ej. fasa recurrente: desarrollo de un mamífero/ movimiento de hojas cíclico; fasa reactiva: variabilidad fenotípica de organismos por el entorno/ los reflejos de los vertebrados; fasa aleatoria: variabilidad fenotípica aleatoria de microorganismos/ la búsqueda del tipo de ambiente que realizan las moscas. Además, también se han presentado criterios experimentales que permiten discriminar cuando un valor de fasa pertenece a un tipo fundamental de fasa.

Una cuestión a considerar sobre la TGE es que no es posible saber si la TGE tiene más leyes que las que han sido formuladas. Además, solo se han formulado postulados y leyes de la TGE mediante las reglas de inducción previamente establecidas. Por lo tanto, aún no se ha desarrollado un análisis matemático de ellas, el cual podría aportar una comprensión más profunda de sus enunciados. Es importante advertir que la TGE no es una explicación a un comportamiento particular, ya que la teoría lo que propone son relaciones generales entre tipos de procesos de computación y las interacciones que se producen entre entidades y entornos. Si se quiere estudiar un exocomportamiento concreto en la TGE, se ha de plantear la ecuación 4.1 donde el exocomportamiento es un valor de la ecuación y la incógnita es una función computacional intensional. La función debe ser intensional para que refleje cada operación computacional, ya que un exocomportamiento no es solo una secuencia de niveles de energía sino también los intervalos de tiempo que duran esos niveles. Por lo tanto, se debe de contemplar toda cuestión que influya en los intervalos de tiempo. Hallándose la función computacional se habría obtenido la descripción de la causa del exocomportamiento. Sin embargo, hay dos problemas en este proceder. El primero es que no existe un método mecánico para resolver la ecuación. El segundo es que pueden existir múltiples soluciones que generen un exocomportamiento. Así, mucha más investigación debe de ser realizada en torno a la ecuación 4.1.

La TGE tiene importantes diferencias con respecto a las propuestas hechas hasta ahora para explicar los comportamientos. Una de las diferencias de la nueva propuesta frente a las teorías unificación de la cognición y sus correspondientes arquitecturas cognitivas [286, 414] es que, mientras que cada arquitectura cognitiva se propone como la única causa por la que sucede el comportamiento cognitivo, la TGE no determina que exista solo una manera de que aparezcan los comportamientos cognitivos. Además, las arquitecturas cognitivas solo consideran procesos computacionales para procesar información, pero la TGE considera que la causa de los comportamientos se identifica con la computación en toda su extensión. Otra diferencia de la TGE comparada con

las teorías matemáticas como AIXI [237], la máquina de Gödel [508], o los ortoretículos [579], es que la TGE es una teoría sobre fenómenos naturales con una causa que es una propiedad física. Por lo tanto, la TGE da unos fundamentos físicos a las ciencias cognitivas, incluida la IA. Respecto a propuestas de carácter físico, como la de Penrose basada en física cuántica [449], o la de Doyle de extender la mecánica clásica [128], resulta que las propuestas mencionadas están ligadas a los conceptos de materia y energía mientras la TGE está conectada a los conceptos de computación e información de física de la información y física fundamental; aunque esto será abordado en el capítulo 5.

Un lector podría preguntarse si es problemático que se deje de usar el concepto de inteligencia para la IA. La respuesta es que no, no lo es. Hechos similares a este han ocurrido ya varias veces en la ciencia. Un ejemplo es el de la química orgánica. Antes del siglo XIX, se creía que los compuestos orgánicos podían ser sintetizados solo por la acción de una “fuerza viva” que solo poseían los organismos vivos. Tras la aparición de la teoría atómica moderna el significado de “compuesto orgánico” es que es una sustancia que contiene una cantidad significativa de carbono. Muchos de los compuestos orgánicos que se conocen hoy no tienen conexión alguna con las sustancias encontradas en los organismos vivos. Un campo científico debe de ser redefinido si esto le permite explicar mejor la naturaleza.

En el comienzo de la IA, Newell and Simon mostraron que los programas informáticos podían funcionar como teorías de los comportamientos por su capacidad de explicar comportamientos [416]. Así, ambos propusieron que la explicación de los comportamientos racionales de los seres humanos podría ser formulada mediante programas de ordenador [414, 532]. La TGE está de acuerdo con la propuesta que hicieron Newell y Simon, porque una parte importante del valor de fase de un sistema exoactivo puede ser descrita por una función computacional, y un programa informático es sólo una de las maneras de describir una función computacional.

Actualmente, las ciencias cognitivas se encuentran en un debate intentando encontrar un criterio para definir cognición [8, 507, 602]. El cognitivismo clásico define un proceso cognitivo como un tipo especial de proceso computacional (simbólico o sub-simbólico) sobre representaciones [602]. Hasta ahora, no hay un criterio de lo que es y lo que no es ese tipo especial de procesos [8, 602]. La TGE considera que ese tipo especial de procesos es simplemente un conjunto de valores de fase. La ciencia cognitiva está enfocada en los procesos cognitivos de alto nivel y son siempre considerados un conjunto aislado de fenómenos. Sin embargo, la TGE no los considera un conjunto aislado de fenómenos sino un subconjunto de un conjunto más amplio de fenómenos. Ante esta situación, uno puede preguntarse que explica mejor el comportamiento la TGE o el cognitivismo clásico. Para dirimir este tipo de cuestiones se suele recurrir al principio de la navaja de Occam. El principio de la navaja de Occam afirma que entre dos explicaciones de un mismo hecho se ha de elegir la explicación que usa menos elementos. Dado el descubrimiento de comportamientos seleccionados por la naturaleza que no son cognitivos, el cognitivismo clásico implica la existencia de conjuntos disjun-

tos de comportamientos y causas distintas para cada uno. Así, siguiendo el principio de la navaja de Occam, la TGE es una mejor explicación de los comportamientos porque la TGE propone la existencia de solo un conjunto y una causa para todos ellos, ya que el cognitivismo clásico requeriría de más elementos para alcanzar la misma capacidad de explicación que la TGE. Además, una teoría del comportamiento debe también explicar los efectos evolutivos de los comportamientos. El cognitivismo clásico no aborda la conexión entre los fenómenos que estudia y fenómenos evolutivos, mientras que la TGE sí lo hace. Pero, estas diferencias no implican que la TGE esté enfrentada con el cognitivismo clásico, ya que las descripciones de procesos computacionales para procesar información para explicar los comportamientos cognitivos son todas directamente interpretables en el marco de la TGE como valores de fase. Así, incluso se podría considerar que la TGE es una extensión del cognitivismo clásico. El problema que emerge es del sector crítico con el propio cognitivismo clásico y que ha expresado que los procesos computacionales son incapaces de explicar el funcionamiento de la mente humana. Esas críticas se abordarán desde el punto de vista de la TGE en el capítulo 8.

## 4.9. Novedades de la reformulación

La formulación que se ha presentado en este capítulo de la TGE tiene ciertas diferencias con las formulaciones presentadas anteriormente en otros documentos[372, 368, 369]. Esas diferencias provienen de cambios que se han hecho para lograr una mejor formulación que evite problemas que se han detectado durante la investigación realizada en esta tesis. Las diferencias que se encuentran en esta formulación son citadas a continuación:

- La primera diferencia proviene del cambio de la definición de exocomportamiento. Mientras que en las anteriores definiciones de exocomportamiento se definía como los estados de las características dinámicas observables y medibles desde el exterior, en esta formulación se define como las cantidades de energía suministradas para lograr el estado que tienen las características dinámicas observables y medibles del sistema. El motivo de acometer en la definición este cambio ha sido el de solucionar el problema de que no se produzca cambio en una propiedad pese a que el sistema esté gastando energía para cambiarlo o a la inversa. Por ejemplo, la posición por la acción de fuerzas exteriores. La nueva definición permite diferenciar entre el comportamiento del sistema cuando la propiedad no está cambiando porque el sistema no gasta energía y cuando la propiedad no cambia pese a que el sistema esté gastando energía. Además, el no diferenciar entre propiedades externas e internas permite que se acoja de manera natural a los fenómenos ontogénicos.
- La segunda diferencia es el cambio en la designación de algunos conceptos. Estos

cambios han sido realizados para evitar connotaciones y definiciones previas que implican ciertos términos en la teoría filosófica del mentalismo. Así, por ejemplo, se ha sustituido fasa sensible por fasa reactiva. También se han cambiado otros términos para que sean más inteligibles. Es el caso de fasa recurrente, la cual fue designada como fasa posicional en anteriores formulaciones. El término recurrente identifica mejor la naturaleza del proceso computacional al que se refiere ese tipo de fasa.

- La tercera diferencia es que en las leyes de fasa se habla de poblaciones de sistemas exoactivos en vez de sistemas exoactivos. El motivo de este cambio es que en el caso de fasa aleatoria la entidad que se selecciona es la población. Así, con motivo de realizar una formulación homogénea en las tres leyes, las tres se han planteado usando a las poblaciones como las entidades que se seleccionan.

## Capítulo 5

# La TGE y su conexión con la física

La física es la base de la ciencia, ya que se ocupa de describir los elementos fundamentales de la naturaleza. Uno de los requisitos que se ha establecido para el objetivo de desarrollar una teoría de unificación de los comportamientos es que sea una teoría fiscalista de los comportamientos. La interpretación que se ha obtenido está basada en el principio de las ciencias cognitivas y de un análisis matemático. Por lo tanto, cabe preguntarse, ¿Hasta qué punto la fasa existe físicamente? Es decir, ¿hay una realidad microscópica con la que se pueda identificar la fasa o se trata de una propiedad que aparece por la escala a la que se observan los sistemas como ocurre con la temperatura en los sistemas macroscópicos?

Así, en este capítulo se va a abordar si la interpretación que se ha hecho de la fasa tiene sentido desde la física. Dado que los valores de fasa son descritos por funciones computacionales, para determinar que la TGE es una teoría fiscalista es necesario poder atribuir significado a esa relación entre los valores y su descripción desde la física fundamental. Para ello, se va a revisar el papel que juegan los conceptos de computación e información en física fundamental.

Desde los años 50 del siglo XX, una nueva visión sobre la naturaleza física del universo ha emergido y ha ido ganando más aceptación. Esta nueva visión considera que los procesos físicos son procesos computacionales y que la información es lo que compone el universo. La nueva visión se ha desarrollado especialmente en las últimas dos décadas [317]. Si la nueva visión de la naturaleza es correcta, la interpretación que se ha hecho de la fasa está totalmente justificada. Hasta ahora, la visión de que el universo es una computadora se ha basado en la conjetura de que la naturaleza del universo es discreta. En contra de la hipótesis de que el universo es una computadora se han propuesto múltiples argumentos. Así, se realizará un análisis de esos argumentos para determinar si realmente contradicen la afirmación de que el universo es una computadora: La estructura del capítulo es la siguiente: la sección 5.1 revisa cómo se ha introducido en la física el concepto de computación y su repercusión. La sección 5.2 revisa cómo se ha introducido el concepto de información en la física y su repercusión.

La sección 5.3 analiza los argumentos en contra de la afirmación de que el universo es una computadora. La sección 5.4 discute cómo afecta las propiedades que tenga la naturaleza en su nivel elemental a la TGE. Finalmente, la sección 5.5 analiza los resultados de la revisión y discusiones que se han llevado a cabo.

## 5.1. Orígenes del concepto de computación y su difusión en la física

El origen de la computación puede ser encontrado en el deseo de los seres humanos de realizar cálculos mecánica o automáticamente. Las razones de querer hacer esos cálculos son muy diferentes. Desde los motivos religiosos de Ramon Llull, o las razones epistémicas de Gottfried Leibniz, a los deseos de resolver problemas matemáticos de Galileo o la exasperación de Charles Babbage debido a los errores en las tablas matemáticas. Durante todo ese tiempo, los trabajos desarrollados sobre cálculos automáticos fueron siempre para llevar a cabo tareas de cálculo. Sin embargo, durante el siglo XX, comenzó una nueva época para la computación, ya que el comprender la propia computación se convirtió en un objetivo. Esta nueva época tiene su comienzo en 1900, cuando David Hilbert comenzó un programa, conocido como formalismo (o el programa de Hilbert), para establecer las bases de las matemáticas y solucionar la profunda crisis en la que estaban sumergidas las matemáticas. Hilbert creía que la manera correcta de desarrollar rigurosamente cualquier materia científica requerías una aproximación axiomática. Sobre esta premisa, él pensó que la fundamentación del análisis requería una axiomatización y una prueba de consistencia. Finalmente, la consistencia de todas las matemáticas podía reducirse a la consistencia de la aritmética básica. Así, afirmó la necesidad de probar la consistencia de los axiomas de la aritmética, y propuso el objetivo de encontrar esa prueba como el segundo de sus 23 famosos problemas matemáticos. Sin embargo, la consistencia no es la única característica que debería ser necesaria para el triunfo del método axiomático. La otra característica es la completitud. Un sistema de axiomas es consistente si no se puede derivar una contradicción de sus axiomas, y es completo si para cada fórmula  $\varphi$  ocurre que,  $\varphi$  o su negación  $\neg\varphi$ , puede ser derivada de los axiomas. Así, los axiomas de la aritmética deberían ser consistentes y completos para tener una sólida fundamentación de las matemáticas. Aunque la meta era clara, el modo válido de probar esas características no estaba claro. En 1922, después de algunas críticas a la propuesta original de Hilbert, este presentó *el punto de vista finitista*. El punto de vista finitista era el modo en el que la consistencia debería ser probada. El punto de vista finitista introduce restricciones que deben regir una demostración de consistencia, y, por lo tanto, en el cálculo para realizar las demostraciones. Dentro de su programa para la fundamentación, Hilbert planteó una pregunta que se conoce como el *Entscheidungsproblem* (el problema de la

decisión)[223]. El *Entscheidungsproblem* pregunta si, dado un sistema de axiomas, existe un método que responda si una fórmula puede ser derivada del conjunto de axiomas o no cumpliendo con el punto de vista finitista. El *Entscheidungsproblem* es el origen de la informática teórica, y específicamente, de la teoría de la computación porque resolver el problema implica definir que es efectivamente calculable, o, en otras palabras, dar un formalismo que permita describir cualquier método para calcular que cumpla el punto de vista finitista.

El *Entscheidungsproblem* está directamente relacionado con la completitud de un sistema de axiomas porque si no hay un método de decisión, entonces el sistema es incompleto. Así, la escuela de Hilbert esperaba que el *Entscheidungsproblem* fuera resoluble. Sin embargo, algo inesperado ocurrió. En 1931, Kurt Gödel demostró dos teoremas que afirman que para cualquier sistema axiomático consistente que incluyera la aritmética de Peano, el sistema contiene proposiciones fuera de su alcance, y por lo tanto, indecidibles. Los resultados de Gödel acabaron con los sueños de Hilbert; pero también cambiaron completamente la visión de los matemáticos y lógicos sobre los cálculos. Antes de los resultados de Gödel, la cuestión de lograr una demostración era una cuestión de habilidad con el cálculo. Sin embargo, los resultados de Gödel mostraron que hay un límite que no puede ser sobrepasado independientemente del nivel de habilidad, mientras se cumpla el punto de vista finitista. Alonzo Church y Alan Turing continuaron el trabajo de Gödel intentando comprender mejor la incompletitud y el poder de cálculo del mecanismo deductivo de los sistemas axiomáticos. Los trabajos de Church y Turing usaron diferentes enfoques. Turing definió una máquina versátil que puede llevar a cabo cualquier computación algorítmica que cumple con el punto de vista finitista. Turing mostró, en consonancia con los resultados de Gödel, que esas máquinas no pueden computar todas las funciones a pesar de su versatilidad, mostrando que *Entscheidungsproblem* es irresoluble [584]. Por otro lado, Church aplicó un enfoque funcional que seguía la línea de Gödel. El enfoque funcional es interesante para evitar las limitaciones de cualquier dispositivo concreto. La cuestión importante de los resultados de Church y Turing es que en ambos casos encuentran el mismo límite.

En un primer momento, uno podría pensar que el resultado de Turing mostrando que hay funciones que no pueden ser calculadas es desalentador. Sin embargo, eso es ver al vaso medio vacío; pero también se puede ver medio lleno, porque si la tesis de Church-Turing mostró que un único mecanismo puede calcular cada función efectivamente calculable, y su número no es pequeño, hay infinitas funciones efectivamente calculables. Este hecho ha sido la clave para la revolución tecnológica en la cual nuestra sociedad está sumergida. Pero la revolución no ha sido solo tecnológica, sino también en nuestra visión de la naturaleza, porque los conceptos de computación comenzaron a impregnar la física. La primera persona que consideró que las afirmaciones computacionales podían tener un significado físico fue Robert Rosen en 1962[491]. Rosen reafirmó la tesis de Church como una afirmación física sobre la no existencia de cierta clase de procesos físicos. Hasta la propuesta de Rosen, las explicaciones de un fenómeno físico se basaban

en postular entidades físicas en la naturaleza; pero Rosen propuso considerar que las limitaciones en computación explican hechos de la naturaleza. Desafortunadamente, sus ideas llegaron demasiado pronto a la comunidad científica porque la cuestión no estaba ni siquiera en consideración.

Uno de los elementos más importante que ha influenciado la investigación de la computación como una cuestión de la física fundamental ha sido el modelo de autómatas celular. La idea de un *autómata celular* (AC) nació como sugerencia de Stanislaw Ulam a John Von Neumann de usar redes de retículos para estudiar el problema de la autoreplicación en los años 40. Norbert Wiener y Arturo Rosenblueth, crearon independientemente de Von Neumann, un modelo de autómatas celular para describir matemáticamente la conducción del impulso en los sistemas cardíacos [621]. Ninguno de los creadores del modelo computacional de autómatas celular tenía en mente la física fundamental cuando lo crearon. Sin embargo, el rol de los autómatas celulares empezó a cambiar cuando Konrad Zuse propuso que el universo entero podía ser computado por un autómatas celular en 1967 [274]. Dos años después publicó su libro *Calculating Space* [275] proponiendo que las leyes de la física son discretas y el universo entero es un AC. La idea de Zuse transformaba completamente el rol de la computación en la física, pero llegó también demasiado pronto para tener impacto en la comunidad científica. La idea tuvo que esperar hasta la existencia de una comunidad de físicos que investigaran los autómatas celulares como herramienta para describir fenómenos físicos. Ese momento llegó cuando Tommaso Toffoli comenzó a estudiar los autómatas celulares en base a la hipótesis de que son modelos naturales para múltiples investigaciones en física [571, 572]. Una vez los autómatas celulares llegaron a ser una herramienta conocida, la comunidad científica empezó a tomar interés sobre el papel que podían jugar en la física fundamental. El interés en los autómatas celulares y la física fundamental explotó finalmente en la conferencia “Physics of Computation” organizada en 1981 en el MIT por Ed Fredkin, Rolf Landauer, y Toffoli donde los asistentes consideraban que física y computación son interdependientes en un nivel fundamental [1]. En la conferencia, entre los asistentes estuvieron Richard Feynman y John Archibald Wheeler. La asistencia de esos dos sobresalientes físicos no fue una casualidad. Las ideas de Feynman podrían ser consideradas como un apoyo al estudio de los autómatas celulares. Muchos años antes de la conferencia, él afirmó:

“ultimately physics will not require a mathematical statement, that in the end the machinery will be revealed, and the laws will turn out simple, like the chequer board with all its apparent complexities” [152].

Wheeler, que fue director de Feynman en su tesis, defendía la necesidad de buscar algo más fundamental que el espacio-tiempo, ya que consideraba que nuestras teorías con el continuo convencional tienen una aplicabilidad limitada. El denominó a ese algo más fundamental, pregeometría [618]. El trabajo de Wheeler enfatizó la idea de

información como un elemento fundamental de la física, pero esto se mencionará más adelante.

La ponencia de Feynman en el congreso fue realmente relevante porque marcó el comienzo de la computación cuántica [151]. La computación cuántica ha sido sin lugar a duda uno de los elementos que ha afianzado la idea de que la computación e información juegan un rol fundamental en la física.

Llegado este punto, es importante notar las dos visiones distintas sobre computación: una es la visión clásica sobre la computación en la naturaleza, en la que los fenómenos físicos son simulados por un modelo computacional [573]. La otra, la visión moderna de la computación, en la que el fenómeno físico es un proceso computacional y las afirmaciones de la teoría de la computación son afirmaciones físicas sobre la naturaleza. Esta visión moderna ha ido tomando más y más relevancia en el campo de la física. En el congreso del MIT de 1981 también participó Zuse que presentó sus ideas. Pero también participo Marvin Minsky presentando un artículo que especulaba sobre considerar el universo un autómata celular y estudiar que física ocurriría en él[387]. Sin embargo, si alguien ha sido un firme proponente junto a Zuse de que el universo es un AC, ese ha sido Edward Fredkin, como el propio Minsky reconocía en su artículo en los agradecimientos.

“This essay exploits many unpublished ideas I got from Edward Fredkin” [387] p.551

Fredkin no escribió sobre sus ideas hasta bastantes años después de que se celebrara el congreso [158]. Sin embargo, él ha sido el promotor de la investigación de los autómatas celulares en la física fundamental, alentando a muchos físicos a hacerlo. Así, hoy se habla de la *tesis Zuse-Fredkin* que se enuncia de la siguiente manera:

*La tesis Zuse-Fredkin: El universo es un autómata celular.*

Así, los autómatas celulares comenzaron a ser un marco para describir procesos al comienzo de los años 80 [626], lo que hizo que algunos físicos comenzaran a preguntarse si los autómatas celulares tendrían las respuestas para ciertas preguntas de la física fundamental [338], y surgieran distintas líneas de investigación. Una de las líneas de investigación más relevantes ha sido la reformulación de la teoría de campos mediante un AC. Entre las investigaciones que se han llevado a cabo, una de las que más ha resonado en la mencionada línea de investigación la realizó Tsung-Dao Lee. Lee propuso que se deberían usar ecuaciones en diferencias para las ecuaciones fundamentales de la física [294, 293, 295]. Sin embargo, sabemos que no todos los AC permiten reformular la teoría de campos. Karl Svozil abordó directamente la cuestión de si los campos cuánticos son autómatas celulares. Svozil demostró que en los autómatas celulares apa-

recía el problema de la duplicación del fermión debido al problema de la imposibilidad [257, 425]. Sin embargo, la generalidad del resultado ha sido criticada por Andrew Ilachinski[240], ya que Svozil no considera una generalización del autómata celular con características cuánticas, y la visión de Svozil está limitada que la de Fredkin[159] o Lee[294]. En 1988, Gerald 't Hooft propuso que la teoría de campos en la escala de Planck podría ser formulada a través de un autómata celular localmente reversible y determinista [233]. Hasta hoy, 't Hooft ha continuado con esa línea de investigación [558]. Así, ha propuesto la interpretación de los autómatas celulares [559]. Por cuestiones de espacio, no se puede revisar la propuesta el trabajo de 't Hooft; pero sus trabajos proponen respuestas a las múltiples preguntas abiertas en la interpretación de Copenhague. La línea de investigación de 't Hooft necesita mucha más investigación porque, como él mismo reconoce, los modelos desarrollados no están suficientemente refinados como para servir de modelos para explicar todas las consecuencias típicas de la mecánica cuántica; pero él ha mostrado que el determinismo no está muerto.

Una nueva línea de investigación surgió en autómatas celulares en 1988 cuando Gerhard Grossing y Anton Zeilinger definió un *autómata celular cuántico* (ACC) como un elemento para estudiar computación cuántica [196]. La idea de un ACC fue propuesta por Feynman en la conferencia *Physics of Computation*, pero Grossing y Zeilinger fueron los primeros en trabajar en ella [195, 197, 165]. En sus investigaciones descubrieron que, en su modelo, excepto para el caso trivial, es imposible una evolución unitaria estrictamente local de toda la rejilla del autómata celular [165]. Por ello, sus modelos son solo aproximaciones cuánticas. Así, David Meyer propuso denominar ACC solo al AC cuya evolución fuese específicamente local, no trivial y unitaria [365, 366]. Meyer y otros autores han usado ACC para estudiar la simulación de retículos cuánticos de gases [365, 60, 320]. En 1995, John Watrous definió otro ACC que resuelve el problema de dar un operador unitario de evolución temporal global [606]. Él definió un ACC bien formado al ACC que su operador de evolución temporal  $T$  es unitario. Watrous dejó pendiente la cuestión de determinar si un ACC es un ACC bien formado o no, pero definió la clase de autómatas celulares cuánticos particionados para los cuales es fácil comprobar si tiene la propiedad de ser bien formado. Poco tiempo después del trabajo de Watrous, se descubrió un algoritmo de tiempo polinómico que comprueba si un ACC está bien formado o no [134, 130]. La investigación del modelo de Watrous mostró la existencia de algunos casos que permiten señales superlumínicas [26]. Así, se han propuesto otros modelos de ACC que evitan señales superlumínica [488, 457, 26]. Recientemente Mauro D'Ariano y sus colaboradores han logrado derivar la teoría cuántica de campos libres a partir de autómatas celulares cuánticos [57, 113].

Debe de notarse que, aunque un ACC podría ser considerado una extensión de un AC, el autómata celular cuántico es una línea de investigación sobre la naturaleza del universo completamente diferente a la del autómata celular. Mientras que el núcleo de la investigación en autómatas celulares es la tesis de Zuse-Fredkin, la línea de autómatas celulares cuánticos considera que el universo debería ser un ACC. La investigación en

la línea del autómata celular propone explicar la rareza del mundo cuántico a través de autómatas celulares que son discretos y deterministas. Sin embargo, la línea del autómata celular cuántico asume que las rarezas del mundo cuántico son propiedades fundamentales de nuestro universo. A pesar de que algunos científicos menosprecian la tesis de Zuse-Fredkin porque involucra determinismo, el trabajo hecho por 't Hooft ha mostrado que es respetable y una línea de investigación [559].

En este punto, es importante que se note que un autómata celular es un mecanismo computacional muy concreto. Sin embargo, independientemente de si el universo es un AC u otro modelo computacional en su nivel más fundamental, hay ciertas limitaciones que debería de haber en la naturaleza dependiendo de a qué clase de sistemas computacionales pertenezca el universo. Los trabajos de Gödel mostraron que hay proposiciones que no son demostrables en ciertos sistemas axiomáticos que cumplan el punto de vista finitista, y Turing demostró por su parte que hay funciones que no son calculables por una computadora cuando la computadora cumple con las restricciones del punto de vista finitista.

La tecnología actual (y no sabemos si la naturaleza también) solo puede calcular funciones dentro del límite de Turing. Así, aparece la pregunta de si las teorías físicas están dentro o sobrepasan el límite de Turing. Esto fue inicialmente planteado e investigado al comienzo de los años 70 por Georg Kreisel [277]. Kreisel propuso diferenciar teorías físicas en base a si eran computables o no por una máquina de Turing. En 1985, David Deutsch, independientemente de la propuesta de Rosen, sugirió que la tesis de Church-Turing debería ser considerada también una afirmación física [124]. La propuesta de Deutsch va más lejos que la de Kreisel porque Deutsch no estaba hablando de una característica de las teorías sino de una característica de la propia naturaleza. Así, Deutsch formuló lo que él denominó el principio de Church-Turing<sup>1</sup> (diferente de la tesis de Church-Turing). Ese principio implica que no existen procesos físicos que estén fuera del límite de Turing. Aunque la propuesta de Deutsch es similar a la propuesta de Rosen, la base de la propuesta de Deutsch es completamente diferente de la de Rosen. Mientras Rosen busca su justificación en la física clásica, los argumentos de Deutsch surgen de la física cuántica. Incluso, Deutsch dio la primera definición de máquina de Turing cuántica para justificar su propuesta. La máquina de Turing cuántica universal ha tenido un rol doble, soporta el principio que enunció Deutsch, pero además abrió la puerta a la tecnología de computación cuántica. Al mismo tiempo que Deutsch hizo su propuesta, Stephen Wolfram, basándose en el teorema de Turing sobre la no existencia de una máquina de Turing que pueda determinar si otra máquina de Turing se detiene o no que muestra irreductibilidad computacional, propuso que la reductibilidad computacional se trataría de la excepción en la física y no de la regla [627]. La existencia de irreductibilidad computacional implica un límite de computación al cuál los sistemas físicos están sujetos.

---

<sup>1</sup>el cual se ha denominado también, el principio de Church-Turing-Deutsch

En los años 90, surgió una nueva conexión entre física y computación. Esta vez fue en el campo de la relatividad general. En 1990, Itamar Pitowsky propuso y desarrolló la idea de usar efectos relativistas para resolver problemas que son irresolubles por una máquina de Turing [464]. Pitowsky definió lo que ahora se conoce como el *espacio-tiempo de Pitowsky*. En 1992, Mark Hogarth generalizó las ideas generales de Malament sobre los requisitos de las estructuras de espacio-tiempo para definir lo que se conoce como el *espacio-tiempo de Malament-Hogarth*, y los aplica a la teoría de la computación [229]. El estudio de la relación entre la relatividad general y computación ha producido el campo teórico denominado computadoras relativistas [429, 144, 137].

Usualmente, la palabra “computable” hace referencia a cualquier cálculo que no rompe el límite de Turing, pero no hay una razón lógica para decir que algo que está fuera del límite de Turing no puede ser físicamente calculable. El propio Turing investigó la idea de dispositivos que pueden superar el límite de la máquina de Turing [585]. El trabajo de Turing comenzó un campo teórico en el que otros muchos investigadores han participado y hecho sus propuestas. En 1999, Brian Copeland y Diane Proud introdujeron el término hipercomputación para cálculos que superan el límite de la máquina de Turing, y que ahora se usa para denominar el campo que inició Turing con sus trabajos. Un buen resumen histórico de las investigaciones del campo de la hipercomputación ha sido realizado por Copeland [106]. Los términos más usados en la literatura para denotar diferentes límites generales de computación son: computable, supercomputable e hipercomputable. Cada uno de esos tres términos implican diferentes características que debería poseer la naturaleza que se manifestarían en los cálculos que pueden llevar a cabo los procesos físicos. El significado de las tres clases anteriores es el siguiente:

- Sistema computacional: El sistema tiene el límite de la máquina de Turing.
- Sistema supercomputacional: El sistema tiene el límite de la máquina de Turing, pero puede resolver problemas de la clase NP en tiempo polinomial, suponiendo que las clases P y NP sean diferentes.
- Sistema hipercomputacional: El sistema puede resolver problemas que no puede una máquina de Turing.

Las tres categorías anteriores se pueden dividir en otras clases más precisas, pero estas tres clases son muy usadas por suponer un gran salto cualitativo entre cada una de ellas. A veces la terminología puede ser un poco ambigua porque en la bibliografía se usa la palabra computación tanto para hablar de cualquier tipo de cálculo, esté dentro o fuera del límite de la máquina de Turing como para referirse únicamente a los cálculos que están dentro del límite de la máquina de Turing.

Cada una de esas clases implica características muy diferentes. Así, la existencia de unos procesos computacionales u otros en la naturaleza sería un indicador de las características del universo (o multiuniverso) en el cual estamos viviendo. El debate de

la hipercomputación es uno de los mejores ejemplos de cómo la computación ha saltado del campo de las matemáticas al campo de la física. Mientras la hipercomputación es sólo un concepto matemático, la existencia de la hipercomputación es un debate físico. Martin Davis ha negado firmemente su existencia [117, 118]. Sin embargo, lo que es verdad sin duda, es que la discusión sobre la existencia de la hipercomputación conduce a verificar conceptos relativistas y cuánticos.

En 2008, Max Tegmark propuso dos hipótesis, la hipótesis del universo matemático (HUM) y la hipótesis del universo computable (HUC) [564]. La HUM aunque ha sido interpretada como una defensa del realismo matemático [244], también puede ser interpretada como una defensa matemática del fisicalismo. Una estructura matemática tiene tres elementos: un conjunto de elementos, un conjunto de relaciones sobre los elementos, y un conjunto de funciones sobre el conjunto de elementos. Como el propio Tegmark ha indicado, cuanto más se desciende en las escalas de la naturaleza, la teoría para describir la teoría se parece más a una estructura matemática y tiene menos equipaje; un ejemplo es la cromodinámica cuántica. Tegmark formuló la HUM de la siguiente manera:

*Mathematical Universe Hypothesis: Our external physical reality is a mathematical structure.*[564]p.102

La definición de estructura matemática dice que contiene un conjunto de elementos, pero eso no quiere decir que los elementos sean entidades supernaturales que determinen un dualismo en la naturaleza. Los elementos pueden ser únicamente elementos físicos. Se puede decir que la HUM es una reformulación de la que ha sido denominada la tesis de Galileo, que él formuló en 1623 de la siguiente manera:

“Philosophy is written in this vast book, which continuously lies open before our eyes (I mean the Universe). But it cannot be understood unless you have first learned the language and recognize the characters in which it is written. It is written in the language of mathematics” [168]

Así, la HUM no es una defensa del realismo matemático sino una defensa matemática del fisicalismo, ya que en ella no hay lugar a la defensa de objetos místicos en nuestro universo. Respecto, la segunda hipótesis de Tegmark, la HUC, determina que la HUM estaría restringida a el conjunto de estructuras matemáticas entre las cuales está la estructura de nuestro universo. La hipótesis ha sido enunciada de la siguiente manera:

*Computable Universe Hypothesis (CUH): The mathematical structure that is our external physical reality is defined by computable functions.*[564]p.131

Independientemente de la propuesta de Tegmark, Matthew P. Szudzik ha propuesto la HUC en 2010 de la siguiente manera:

*Computable Universe Hypothesis. The universe has a recursive set of states  $U$ . For each observable quantity, there is a total recursive function  $\phi$ .  $\phi(s)$  is the value of that observable quantity when the universe is in state  $s$ .*[556]

Actualmente, recursividad y computable<sup>2</sup> son sinónimos. Así, las propuestas de Tegmark y Szudzik son la misma. La HUC es una hipótesis importante para la búsqueda de una teoría del todo. Si la HUC es verdadera entonces sería un criterio para buscar una teoría del todo. Se sabe que la física clásica no es una teoría del todo porque sus predicciones son muy diferentes de lo que realmente ocurre cuando se realizan los experimentos tanto en grandes como en las pequeñas escalas de la naturaleza. Sin embargo, si se acepta la HUC se puede llegar a la misma conclusión. Existen múltiples resultados teóricos que muestran como la física clásica permite superar el límite de la máquina de Turing. Por ejemplo, se ha demostrado que hay soluciones para la ecuación de ondas que no son computables [470, 469], y resultados directos que muestran cómo resolver problemas indecidibles e incompletos en la mecánica clásica [108]. Por lo tanto, si la HUC es verdad, la mecánica clásica no puede ser una teoría del todo porque los resultados de incomputabilidad sobre la mecánica clásica, lo que está en consonancia con lo que los experimentos nos han mostrado, ya que el mundo cuántico y los sistemas con una alta gravedad no pueden ser explicados por la mecánica clásica.

Una cuestión a tener en cuenta es que, si se demostrara que la hipercomputación existe, la HUC y el principio de Church-Turing serían falsos. Además, la HUM no implica la tesis de Zuse-Fredkin porque la HUC no propone el mecanismo que gobierna la evolución del universo, sino una característica del mecanismo que gobierna la evolución del universo. Sin embargo, la tesis de Zuse-Fredkin si implica la HUC porque un autómata celular no excede el límite de Turing.

La última relación que se ha encontrado entre física y computación ha surgido en las investigaciones teóricas de los agujeros negros. Concretamente, Daniel Harlow y Patrick Hayden han propuesto que la solución a los muros de fuego en el horizonte de sucesos podría estar relacionada con la complejidad computacional de los procesos que sucederían. [207].

---

<sup>2</sup>en el sentido de los procesos dentro del límite de la máquina de Turing

## 5.2. Orígenes del concepto de información y su propagación en la física

En física, nadie estaba preocupado por la información hasta la segunda mitad del siglo XX. Sin embargo, hoy en día es uno de los principales elementos en las discusiones de la física. Para entender este gran cambio, hay que retrotraerse al siglo XIX, ya que las matemáticas de la teoría de la información están implícitas en los trabajos de Ludwig Boltzmann de 1872. Sin embargo, el concepto de información no surgió explícitamente hasta los trabajos de Ronald Fisher en el campo de la estadística. Fisher definió en 1925 un método para medir la cantidad de información aportada por datos de un parámetro desconocido. En 1928 Ralph Hartley definió el logaritmo del número de los símbolos posible en una secuencia como una medida de información [210]. En 1929 Leo Szilard escribió un artículo sobre el demonio de Maxwell que abrió la puerta de la física a la información [555]. En el artículo decía lo siguiente:

“A perpetual motion machine is possible if -according to the general method of physics- we view the experimenting man as a sort of *deus ex machina*, one who is continuously informed of the existing state of nature...”

Así, por primera vez, la idea de la información aparecía en la física. Aunque Szilard no dio una medida general de la información, conectó la idea de medir el estado del entorno y la generación de entropía a partir de la formulación de Boltzmann.

A pesar de todo esto, no fue hasta los trabajos independientes de Claude E. Shannon [524, 525] y Norbert Wiener en 1948 cuando la teoría de la información <sup>3</sup> ganó renombre debido a la aplicabilidad en el campo de la ingeniería de comunicaciones. La formulación matemática de la teoría de información de Shannon y la teoría de Boltzmann eran tan similares que, en un encuentro con John von Neumann, Shannon le preguntó su opinión sobre cómo llamaría él a la medida de incertidumbre de su teoría, a lo que von Neumann le respondió:

“You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage” [578]

Debe de notarse que la entropía de la información de Shannon es un concepto mucho más general que la entropía de la termodinámica estadística. La entropía estadística de la información es un caso específico de la entropía de la información porque la entropía

---

<sup>3</sup>La noción de información de Fisher y Shannon son diferentes

de la información está presente siempre que hay cantidades desconocidas que pueden ser descritas solo por una distribución de probabilidad.

Poco después de los trabajos de Shannon y Wiener, diversos físicos notaron la importancia de la teoría de la información para interpretar la termodinámica. Uno de ellos fue Leon Brillouin, que introdujo la teoría de la información en el debate del demonio de Maxwell a través de la puerta de Szilard [77]. Brillouin identificó el concepto de entropía negativa, o neguentropía, propuesto por Schrödinger [513], algunos años antes de la aparición de la entropía de Shannon. Eso supuso el salto del concepto de la información a la termodinámica. Desde entonces, la teoría de la información ha sido un elemento clave en las discusiones del demonio de Maxwell. En 1953, Brillouin escribió una ecuación fundamental afirmando que el cambio de un bit de información requiere al menos  $kT \ln(2)$  de energía. Esa es la misma energía que el motor de Szilard produciría en el caso ideal. En 1957, Edwin T. Jaynes propuso que la mecánica estadística puede ser fundamentada usando la teoría de información [245, 246]. Jaynes formuló el principio de máxima entropía, y lo propuso como la base sobre la que desarrollar la mecánica estadística. Así, la noción de información comenzó a ser un elemento común en termodinámica.

El siguiente gran hito en el uso del concepto de información en la física fue realizado por Jacob D. Bekenstein 1973. Bekenstein aplicó la teoría de la información con los agujeros negros [48, 49], permitiéndole demostrar que los agujeros negros tienen entropía. Él introdujo la entropía como una medida de la información sobre el interior de los agujeros negros y determinó que el área de su horizonte de sucesos debe ser proporcional a su entropía. A continuación, Stephen Hawking demostró que un agujero negro debe emitir partículas y calculó su temperatura [212]. Tras ese brillante descubrimiento, Hawking propuso que los agujeros negros destruyen información [211]. Pero, ¿a qué información se refiere esa afirmación? Para entender la afirmación de Hawking se tienen que considerar dos hechos. Primero, cualquier estado físico almacena información. Segundo, las leyes de la física clásica y cuántica son reversibles, y, por lo tanto, mantienen la distinción entre estados. Si la distinción no se mantuviera habría información que sería destruida. Sin embargo, las leyes de la física clásica y cuántica no destruyen información. Ese hecho, también apuntaría a que una condición necesaria de una teoría física es que conserve la información. En el caso de la física clásica, el teorema de Liouville asegura que la información se conserve. En la mecánica cuántica, la restricción de unitariedad juega el mismo rol que el teorema de Liouville tiene en la mecánica clásica. Según Hawking, al combinar mecánica cuántica y relatividad general había información que se destruía, pero esto iría en contra de las leyes de la física cuántica. Esto es lo que se ha conocido como la paradoja de la información, y abrió la puerta de la física fundamental al concepto de información. Entre los intentos para resolver la paradoja, una propuesta es considerada actualmente la más probable, el principio holográfico. El principio holográfico fue propuesto independientemente propuesto por Gerard 't Hooft, y por Leonard Susskind. El principio holográfico propone que la realidad que se ob-

serva sería como un holograma generado por la información que se encuentra en otra superficie. La primera formulación matemática del principio holográfico fue dada por 't Hooft en 1993 [557]. En 1995, Susskind dio una precisa interpretación del principio en la teoría de cuerdas y propuso la existencia de un límite en la cantidad de información que puede contener un sistema si se cumple el principio holográfico [554]. Más tarde, Bekenstein mostró cómo ese límite se podía derivar a partir de una generalización de la segunda ley de la termodinámica [50]. En la actualidad, el principio holográfico es uno de los temas habituales en la investigación de la física teórica.

Además de la paradoja de la información, hay otra paradoja por la que el concepto de información ha tomado relevancia en la física, la paradoja de Einstein-Podolsky-Rosen(EPR). En la interpretación del experimento de Einstein-Podolsky-Rosen se proponía que el entrelazamiento cuántico estaba en conflicto con la teoría de la relatividad. Para demostrar que no hay tal conflicto, se ha probado que no es posible enviar información más rápido que la velocidad de la luz en el vacío a través del fenómeno de entrelazamiento cuántico [180, 181].

En la revisión que se ha hecho, se puede ver cuánto ha cambiado el punto de vista en la física sobre la información. Casi 70 años atrás, la información era considerada una noción abstracta que debía ser implementada mediante sistemas físicos. Hoy en día, la información es considerada una noción fundamental de la naturaleza. En este cambio, sin duda se debe mencionar a John A. Wheeler ya que inspiró a muchos físicos a usar el concepto de información. Wheeler propuso y defendió que la información es la base del universo. Esa propuesta es reflejada muchas veces en su famosa cita “*It from the bit*” [619]. En consonancia con la propuesta de Wheeler, recientemente se ha derivado la física cuántica de la teoría de la información [99]. La situación es que el concepto de información resulta clave en la física fundamental. En 2003, Bekenstein ya afirmó que el objetivo por el que estaba creciendo el interés era el de definir cómo el mundo físico se construye a partir de información y que había que abandonar la idea de que el lenguaje de la teoría de campos es el lenguaje final de la física[47].

### 5.3. Argumentos en contra de que el universo sea una computadora

En esta sección se va a revisar y discutir seis argumentos que se han propuesto en contra de la propuesta de que el universo es una computadora. La afirmación de que el universo es una computadora sería falsa si existiera un proceso físico que no fuera computacional. Los contraargumentos que se han propuesto tratan de demostrar que hay procesos que no son computacionales. Sin embargo, se va a mostrar que ninguno de los contraargumentos es válido porque en todos se identifica computación con una subextensión de la extensión real que posee el concepto de computación.

### 5.3.1. El contraejemplo del sistema radioactivo

El primer argumento en contra de la afirmación de que el universo es una computadora fue propuesto incluso antes de que la idea naciera. El contraargumento está en el artículo de 1962 de Rosen[491]. Rosen menciona que George Yuri Rainich propuso que hay un ejemplo de sistema físico que no puede ser descrito computacionalmente que consiste en un sistema que contiene material radiactivo de manera que la salida del sistema es consecuencia de la emisión de partículas del material radiactivo. Puesto que la emisión de las partículas por la radioactividad se hace de manera aleatoria, el sistema sería no describible computacionalmente. Rosen intento abordar el problema sugiriendo que la función del sistema radiactivo sería no Turing-computable. Sin embargo, la propuesta de Rosen no es la mejor manera de abordar el problema. Cuando Rainich sugirió su ejemplo, la computación probabilística no había nacido. La definición de autómeta probabilista fue dada en 1963 por Michael O. Rabin [475], y la definición de máquina de Turing probabilista fue dada en 1969 por Eugene S. Santos[502]. Obviamente, un sistema radioactivo no puede ser descrito por una máquina de Turing porque es un sistema determinístico; pero una máquina de Turing probabilista puede hacerlo. Así, el contraargumento del sistema radiactivo es incorrecto porque está considerando que los procesos computacionales se limitan a los procesos determinísticos; pero el conjunto de los procesos computacionales es en estos momentos más amplio que los determinísticos, ya que también están los procesos computacionales probabilísticos.

### 5.3.2. El contraargumento del continuo

Otro contraargumento que ha sido propuesto contra la afirmación de que el universo es una computadora es el del continuo. David Tong ha atacado la idea de que la computación juega un papel fundamental en la naturaleza porque él defiende que la naturaleza es continua [574]. Tong argumenta que no hay discretitud en la ecuación de Schrodinger, sino que lo discreto es generado por la ecuación, y que las partículas fundamentales son rizados de campos continuos. Así que, de acuerdo con eso, la discretitud de las partículas no es una característica fundamental de la naturaleza, sino que emerge. El objetivo de este argumento es rechazar la idea de que la discretitud que se ha encontrado en la naturaleza es una similitud no casual con las computadoras digitales. El debate de si la naturaleza es continua o discreta comenzó a principios del siglo XX con el nacimiento de la física cuántica. Quizás Tong ha considerado que se podía reformular en términos computacionales, como la discusión entre procesos computacionales y no computacionales considerando que computación implica discretitud. Determinar si la naturaleza es discreta o continua está completamente fuera de los objetivos de esta tesis; pero hay un problema con el contraargumento porque la reformulación del debate es incorrecta. El origen de identificar computación con discretitud viene del hecho de que efectivamente calculable implica discretitud por el punto de vista finitista. Así,

muchos científicos consideran que el poder computacional del mecanismo que gobierna el universo está dentro del límite de la máquina de Turing, y la discretitud es una de sus propiedades [125]. Como se ha visto en la sección 5.1, el desarrollo de la visión de que los procesos fundamentales de la naturaleza son computacionales se ha debido a la discretitud encontrada en la naturaleza. Sin embargo, ni computación ni información implican la discretitud. Tal vez se identifique computación con discretitud porque en el mercado se encuentran prácticamente solo computadoras digitales de propósito general y se ha identificado computadoras con tecnología, pero en los años 60 había computadoras analógicas comerciales [243]. El primer modelo teórico de computadora analógica fue realizado por Claude E. Shannon en 1941 [526]. Sin embargo, durante mucho tiempo han faltado unos fundamentos teóricos coherentes para las computadoras analógicas, lo que ha impedido su mejora. Eso ha cambiado durante las últimas décadas. Christopher Moore dio en 1996 un nuevo enfoque que ha sido exitoso creando una base teórica. Moore definió un conjunto de funciones sobre los reales de un modo similar al método de las funciones recursivas sobre los números naturales [400]. Esa clase de funciones recursivas es denominada funciones  $R$ -recursivas. Este enfoque ha dado lugar a nuevos resultados [69, 107, 67, 410]. Además, se han creado versiones para dominios continuos de  $\lambda$ -Calculus [2, 562], y más recientemente se ha dado una definición de algoritmo que captura el comportamiento [68].

En base a todo lo anterior es claro que los conceptos de computación e información son definibles tanto en dominios discretos como continuos. Así que ni información ni computación implican discretitud, y que la naturaleza fuera continua es su nivel más fundamental no invalida la afirmación de que el universo es una computadora.

### 5.3.3. El contraejemplo de la evaluación

Giusseppe Longo y Thierry Paul han analizado la cuestión de si un proceso físico lleva a cabo computación [166]. Su opinión es que están en contra de considerar que un proceso físico computa. En su razonamiento, identifica computación con ecuaciones y considera que la resolución de ecuaciones es solo uno de los métodos para describir sistemas dinámicos en física. Longo y Paul resaltan en su argumentación que la evaluación es un método diferente del de la resolución de ecuaciones, y muy importante, ya que está detrás del principio de mínima acción y la integral de caminos. Así, afirman:

*“ This principle of least action does not ask to solve an equation, it just asks to evaluate the functional  $S$  at any possible path  $\gamma$  and select the extremal one.”* [166] pag.252

*“But the Feynman ‘path integral’ formulation of quantum mechanics create a revival of this idea of evaluating instead of computing.”* [166] pag.252

Esta última afirmación podría ser considerada más importante ahora, tras un reciente experimento que apoya la formulación de la integral de caminos[330].

Sin embargo, el contraargumento está equivocado. Longo y Paul distinguen las operaciones de resolver ecuaciones de las operaciones de evaluar y seleccionar; pero todas las operaciones son computacionales. Las operaciones para resolver ecuaciones son reglas de reescritura que producen un proceso de reducción que termina, y en el que las formas normales son únicas. Evaluar es también un tipo de proceso realizable por reglas de reescritura con las características de terminación y confluencia. Seleccionar un punto estacionario es un proceso de reescritura, pero además es una operación computacional muy importante llevada a cabo por el operador de minimización en la teoría de funciones recursivas. El operador de minimización,  $\mu$ , puede ser definido tanto en los enteros como en los reales [400]. La teoría de los sistemas de reescritura es un marco para describir procesos computacionales. Cualquier máquina de Turing puede ser traducida a un sistema de reescritura de términos finitos. Un sistema de reescritura puede ser visto como un programa de computadora cuando es combinado con un algoritmo apropiado, y múltiples lenguajes de programación declarativa en la reescritura de términos[30].

### 5.3.4. El argumento de las infinitas dimensiones

Además del anterior argumento, Longo y Thierry Paul propusieron otro contraargumento para no considerar que los procesos físicos computan[166]. Ellos argumentan que en mecánica cuántica es necesario un espacio de Hilbert con infinitas dimensiones para representar observables continuos como la posición o el momento angular. Esa necesidad se debe a que en un espacio con un número finito de dimensiones, la relación de conmutación canónica  $[x, p] = i\hbar 1$  no se puede llevar a cabo, porque con finitas dimensiones  $x$  y  $p$  son operadores de clase traza, y la traza de sus conmutadores se desvanece. De ese hecho se concluye que se necesita un espacio para el cual la traza de la identidad no esté bien definida, lo que ocurre cuando el espacio tiene infinitas dimensiones[616]. Longo y Paul consideran que este hecho muestra que el marco de computación no es adecuado para describir el universo. Aunque algunas propuestas alternativas han sido hechas para dar un tratamiento finito al problema [500, 501], de nuevo, no es necesario tomar parte en la discusión física de la cuestión porque se puede mostrar que la necesidad de infinitas dimensiones no implica que un proceso suceda en un espacio de infinitas dimensiones determine que no sea computacional. El motivo por el que el contraargumento de las infinitas dimensiones es incorrecto es que existen modelos computacionales que permiten considerar esos procesos como computacionales[612, 614, 613].

### 5.3.5. El argumento del esquema lagrangiano

Otro contraargumento que se ha esgrimido frente a la afirmación de que el universo es una computadora ha sido propuesto por Ken Wharton [617]. Wharton ha argumentado que la computación es un proceso que solo puede existir en el esquema newtoniano. Wharton define el esquema newtoniano como el método en el que primero se mapea el mundo físico en algún estado matemático, después se usan las leyes dinámicas para transformar ese estado en otro nuevo estado que puede ser interpretado también en el mundo físico. Wharton explica que, aunque la mayoría de las teorías físicas usan el esquema newtoniano, este no es el único mecanismo con el que plantear una teoría física, y nuestro universo podría no ser un universo de esquema newtoniano. Wharton propone que al menos hay un mecanismo alternativo al esquema newtoniano, el esquema lagrangiano. El esquema lagrangiano está inspirado por la mecánica lagrangiana. El esquema lagrangiano mapea el mundo físico en un espacio de estados, y el estado inicial y final son determinados en el espacio de estados que son los parámetros que están restringidos. El camino entre el estado inicial y final es el parámetro libre que debe ser generado a través de una regla global. El argumento de Wharton tiene tres puntos. Primero, el esquema newtoniano es desafiado por la naturaleza cuántica del mundo. Segundo, el esquema lagrangiano aborda mejor los desafíos de la naturaleza cuántica del mundo. El punto final es que él argumenta que si el universo es un universo de esquema lagrangiano entonces el universo no es una computadora porque no calcula un programa imperativo. El contraargumento de Wharton puede parecer que es el mismo que el del apartado 5.3.3, pero no lo es. Mientras que el argumento de Longo y Paul está relacionado con la naturaleza matemática de las operaciones, Wharton señala con su argumento a la evolución del universo como la cuestión clave.

De nuevo, no se va a entrar a discutir si el esquema lagrangiano o el esquema newtoniano pueden abordar mejor la naturaleza cuántica del mundo, pero la afirmación de que el esquema lagrangiano niega que el universo sea una computadora es errónea. La razón de que sea errónea es que la afirmación de Wharton de que solo las teorías que pertenecen al esquema newtoniano son computacionales es falsa. El esquema newtoniano puede ser identificado con el programa imperativo pero el paradigma imperativo no es el único. Hay procesos computacionales que pertenecen a otros paradigmas computacionales. Entre los paradigmas computacionales que existen, uno de ellos es el de la programación lógica [24]. En programación lógica las variables son valores desconocidos. Un programa escrito en un lenguaje de programación lógica es un conjunto de sentencias en forma lógica, expresando hechos y reglas sobre el dominio del problema. Los lenguajes de programación lógica son diferentes de los lenguajes de programación imperativa. En los programas lógicos no hay una contrapartida al concepto de incremento de un valor. De la misma manera que el principio de Fermat, en programación lógica se da el estado inicial y final y una regla global en el programa que permitirá encontrar los valores que responder la consulta. El proceso del esquema lagrangiano

propuesto por Wharton es similar a los programas lógicos. El lenguaje de programación Prolog pertenece al paradigma lógico. Prolog es usado alrededor del mundo entero para crear programas informáticos. Por lo tanto, el contraargumento es incorrecto y no puede rebatir la afirmación de que el universo no es una computadora.

### 5.3.6. El contraejemplo de los oráculos

Tarner Edis y Maarten Boundry han propuesto una solución al dilema de Hempel [140]. El dilema de Hempel es la pregunta de cómo decidir qué es natural y qué es sobrenatural. Edis y Boundry han propuesto que algunos procesos computacionales son naturales y que otros procesos computacionales serían sobrenaturales lo que permitiría determinar los límites de la física. A diferencia de los contraargumentos anteriores que consideran que la computación no es suficientemente general para ser considerada una propiedad física universal, el argumento de Edis y Boundry implica que el concepto de computación sería demasiado general porque permitiría describir procesos sobrenaturales. Así, una teoría computacional no sería fiscalista porque no sería monista sino dualista. En su razonamiento, ellos afirman:

“*When doing physics, we only have access to finite computational resources*” [140]p.407.

Edis y Boundry aseguran que hacen esa afirmación porque no es controvertida. Sin embargo, esa afirmación es completamente controvertida [200]. No se sabe si el espacio o el tiempo son discretos o continuos, y es una de las cuestiones más importantes sin respuesta que existe actualmente en la física. Un experimento reciente realizado en la escala de Plank no ha observado correlaciones exóticas en sus medidas del espacio-tiempo [101]. Esos resultados han excluido una teoría de un universo holográfico con un alto nivel de significancia estadística. Si las fluctuaciones cuánticas existen, son mucho más pequeñas que la escala de Plank. En el futuro se realizarán muchos más experimentos para comprobar si el espacio-tiempo es continuo o no. En esta tesis no se toma partido en la discusión física sobre si el espacio-tiempo es continuo o no, pero es claro que la afirmación de la que parten es controvertida. Además de lo anterior, el punto importante en el contraejemplo de Edis y Boundry son los oráculos significativos. Significativo hace referencia a que los datos de salida del oráculo son las respuestas correctas a un problema que no puede ser resuelto por una máquina de Turing. Cuando Turing propuso la idea de oráculo, él consideraba que era una caja negra que produce como salida a un valor de entrada, el valor de una función incomputable por una máquina de Turing en un solo paso. Usualmente, en la teoría de la computación, los oráculos son estudiados desde el punto de vista de la complejidad de los conjuntos que pueden definir. Esta parte de la teoría de la computación se denomina *computación relativa*. La computación relativa es el apartado más teórico de la teoría de la computación. Pero, por otro lado, es interesante notar que un proceso hipercomputacional no es algo

extraño en física. Una integral realiza un cálculo con un número infinito de datos para generar un dato de salida. Una simple integral definida como la siguiente:

$$\int_0^1 x dx$$

es un cálculo que no puede ser llevado a cabo por una máquina de Turing porque no puede efectuar un número infinito de sumas para alcanzar un estado final en una cantidad finita de tiempo. Sin embargo, usando el teorema fundamental del cálculo se puede calcular el resultado de la integral en un número finito de pasos en una cantidad finita de tiempo. Así, se sabe que:

$$\int_0^1 x dx = 0'5$$

Cualquier objeto moviéndose con un movimiento rectilíneo está llevando a cabo los cálculos de una integral según la mecánica newtoniana.

Por otro lado, si el universo tuviera una cantidad finita de recursos no habría manera de comprobar que un objeto es un oráculo, porque la cuestión clave de un oráculo no es que decida que una máquina de Turing se detiene o no con unos datos de entrada, sino que puede hacerlo con todas las posibles entradas, y eso significa infinitas entradas, aunque cada una sea finita. Por lo tanto, en un universo con recursos finitos hay un número máximo de entradas con las que se pueden dar como datos de entrada al supuesto oráculo.

No se puede decir si computación es el marco final para abordar los límites del fisicalismo porque nadie sabe si en 100 años alguien propondrá una teoría rompedora que deje obsoletos los conceptos computacionales; pero sin duda los conceptos computacionales están dentro del fisicalismo. Descubrir un oráculo en un universo finito es una contradicción. Eso involucraría que el objeto tiene mucha más información que lo que las leyes físicas permitirían a un sistema físico. La afirmación de encontrar un oráculo en un universo finito es similar a la afirmación de descubrir una persona que meditando puede levitar en un universo en el que las leyes no permiten levitar. Obviamente, ambas implican objetos sobrenaturales porque si se definen los límites de lo que puede ocurrir, descubrir algo que está fuera de los límites implica que es un elemento sobrenatural. Este razonamiento funciona con computación y cualquier otro marco fiscalista que se plantee. Sin embargo, hay un problema con el razonamiento. El problema es que Edis y Boundry asumen que las leyes finales del universo son conocidas, pero no lo son. Las leyes que se tienen se han creado por inducción y no tienen porque ser correctas. Edis y Boundry están considerando que el resultado de un experimento físico tiene que ser acomodado a las leyes que se tiene del universo, y no se dan cuenta que los experimentos son lo que determinan si un principio o teoría es correcta o equivocada. El razonamiento de Edis y Boundry va en contra del principio de la navaja de Occam.

Ellos proponen que si se encuentra un oráculo significativo habría que considerar que existen dos tipos de objetos, naturales y sobrenaturales, pero eso implica multiplicar las entidades. Según la navaja de Occam, hay que escoger la explicación que tiene menos entidades, que en este caso es que nos encontramos en un universo cuyas leyes permiten la hipercomputación. En este sentido, existen múltiples trabajos teóricos que estudian la computación con oráculos que son sistemas físicos [41, 43, 42, 44].

## 5.4. Propiedades fundamentales de la naturaleza y la TGE

En la sección 5.1 el debate sobre si la naturaleza es un su nivel más fundamental es determinista o aleatoria [559]. Así, pueden surgir las siguientes preguntas respecto a la física: ¿cómo afectaría a la TGE si se demostrase que el universo es determinista o aleatorio en su nivel más elemental? y ¿quedarían invalidados el principio del carácter de los valores de fase y el teorema fundamental de la fase?

La respuesta es que no tendría ningún efecto sobre la TGE. La razón es que un exocomportamiento requiere la existencia de un sistema físico con frontera. Eso quiere decir que es un fenómeno que no está ocurriendo en la escala fundamental. En el caso de que se demostrase que en el nivel más fundamental de la naturaleza los fenómenos son aleatorios, desde el punto de vista computacional un sistema determinista es un caso especial de sistema aleatorio donde cada estado tiene siempre una con probabilidad 1 y el resto con probabilidad 0. Eso quiere decir que si se anula la probabilidad de ir hacia otros estados se logra un sistema determinista. La situación sería semejante a lo que ocurre en el mecanismo del flagelo bacteriano que es capaz de generar un movimiento en una dirección a pesar de encontrarse en un medio de movimiento browniano. [334].

En el caso de que se demostrará que la naturaleza es determinista en su nivel más fundamental tampoco esto afectaría a la TGE porque la fase aleatoria seguiría existiendo. Para ellos es suficiente que los sistemas de una población tengan generadores de números pseudoaleatorios con semillas distintas que provoquen que los exocomportamientos varíen aleatoriamente. La obtención de una semilla diferente no es una cuestión problemática, ya que incluso se podría obtener en el nivel más fundamental de la naturaleza. En la interpretación del autómata celular de la física cuántica propuesta por t' Hoopf, para explicar la aleatoriedad que se observa pese a que su teoría es determinista, propone que las fluctuaciones del vacío, que son diferentes en cada situación son la fuente de la aparente aleatoriedad que se observa en las partículas ya que es una variable que no puede ser medida para ser incluida en los cálculos [559].

Otra cuestión a resaltar sobre la TGE tras la revisión de este capítulo es que el poder computacional de los valores de fase viene determinado por el límite computacional del universo. La TGE no está restringido al límite de la máquina de Turing ni obliga a

la existencia de valores de fasa hipercomputacionales. Los límites computacionales de la TGE solo son determinado por el límite computacional de los procesos físicos del universo.

## 5.5. Análisis de los resultados de la revisión

La revisión que se ha realizado ha mostrado que los conceptos de computación e información son actualmente muy importantes para entender el nivel más fundamental de la naturaleza y que todo apunta a que su importancia se incrementará aún más en las próximas décadas. Esto implica que la interpretación de la fasa como una propiedad computacional que se ha desarrollado y formulado en la TGE es consistente con la física que se conoce a día de hoy. Además, sean cuales sean las características que posea el nivel más fundamental de la naturaleza, ni el principio del carácter de los valores de fasa ni el teorema fundamental de la fasa dependen de esas características.

Un lector podría estar preocupado por la idea de que solo una parte de la computación que se lleva a cabo en el sistema exoactivo sea la causa de los exocomportamientos. Sin embargo, el hecho de que el universo fuera una computadora no colisiona con la idea de que la fasa sólo sea una parte de la computación que ocurre en el sistema. Considérese como se aborda actualmente la descripción de un sistema físico mediante energía. El sistema tiene una energía total, pero la energía total es la suma de diferentes tipos de energía: la energía térmica, la energía cinética y la energía de los enlaces químicos entre otras. Cada una de esas energías está envuelta en un tipo diferente de fenómenos. Esa misma situación sucede en los sistemas exoactivos. Un sistema exoactivo puede verse como un sistema donde todo es un proceso de computación, pero solo un determinado proceso de computación llevado a cabo por la fasa es la causa directa de los exocomportamientos.



## Capítulo 6

# La fasa y el sistema nervioso I: Dinámicas cerebrales

En el capítulo 4, al hacer el análisis matemático de la fasa se consideró  $M$  desde un punto de vista funcional. Sin embargo, como ya se dijo, el valor de fasa  $\mathcal{F}$  es conformado por el modelo computacional, los datos que posee y la implementación física del modelo computacional. Además, se ha mencionado que el valor de fasa de sistemas exoactivos es determinado por procesos evolutivos cuando los sistemas exoactivos son unidades de evolución. Por lo tanto, esto lleva a considerar que el modelo computacional, los datos que contiene y la implementación física del valor de fasa de un sistema exoactivo que es una unidad evolutiva debe ser explicable por procesos evolutivos. Así, en este capítulo se abordan tres cuestiones respecto al valor de fasa de los animales: ¿cómo representan estados los sistemas nerviosos de los animales?, ¿qué proceso evolutivo ha determinado la representación de información en las redes neuronales biológicas? y ¿cuál es el origen evolutivo del modelo de red neuronal en el sistema nervioso? El autor no ha encontrado trabajos en los que la segunda y tercera cuestión hallan sido planteadas en la literatura del campo de la evolución del sistema nervioso antes de esta tesis doctoral. Este capítulo va a tratar la primera y segunda pregunta. La tercera pregunta será abordada en el siguiente capítulo.

La actividad cerebral se mide a diferentes escalas usando diferentes técnicas EEG, MEG y fMRI [206, 588]. Esas técnicas están relacionadas con la sincronía de un gran número de neuronas. Por lo tanto, aunque EEG, MEG y fMRI son útiles para investigar el comportamiento cognitivo, no permiten descubrir o investigar el mecanismo neuronal que permite el flujo de información ni explicar cómo se procesa la información. Desde la segunda mitad del siglo XX, la electrofisiología ha investigado en profundidad como procesa señales una neurona mediante microelectrodos. Así, las características funcionales de las unidades de procesamiento del cerebro se conocen muy bien desde finales de los años 80 [312]. Sin embargo, existen diversas investigaciones teóricas y experimentales que han indicado que la información puede estar codificada en patrones

de actividad dispersos [367]. Por lo tanto, si se desea entender la función de una población de neuronas y se intenta inferir la función a partir de registros independientes de neuronas individuales, sucederá que se pasará por alto aspectos importantes de la información codificada en patrones de cambio de la actividad que están distribuidos a través de la población de neuronas.

Durante largo tiempo, no existían técnicas para realizar registro de cada neurona de una población simultáneamente. Sin embargo, para rellenar el hueco entre la actividad de una neurona y la del cerebro, se han propuesto e investigado modelos matemáticos sobre los mecanismos que las poblaciones de neuronas del cerebro podrían usar para cumplir su función. Concretamente, se han formulado dos posibles tipos de mecanismos de neuronas subyacentes a las dinámicas neuronales: vector de estado e itinerancia caótica. Cada uno de estos tipos de mecanismos agrupa diversos métodos. En el siglo XXI, la discusión ha empezado a pasar del plano teórico al experimental, ya que ahora es posible registrar simultáneamente múltiples neuronas de una población. Una de las técnicas que permite eso es el método de imagen de calcio [551]. El método de imagen de calcio permite estudios de actividad neuronal en cientos de neuronas dentro de circuitos neuronales. Estos avances han permitido recientemente la observación de dinámicas neuronales [38]. La investigación de Brice Bathellier y sus colaboradores [38] ha mostrado que las capas superficiales del córtex auditivo contienen un número pequeño de grupos de neuronas actúan como atractores que producen la categorización de sonidos. La cuestión de cómo las dinámicas del cerebro operan parece estar alcanzando una nueva fase donde los experimentos pueden dar información de cómo funciona. Pero a pesar de los esfuerzos de investigación sobre cómo los sistemas nerviosos codifican información, solo se tienen propuestas teóricas con poca validación experimental. Aún así, la discusión de las propuestas teóricas es importante porque cada propuesta esta relacionada con conjuntos de valores de fasa, entre los que se encuentran los de los seres humanos.

La segunda pregunta de ¿qué proceso evolutivo ha determinado la representación de información en las redes neuronales biológicas? sólo es posible plantearla, ya que hallar una respuesta a ella depende completamente de tener una respuesta a la primera pregunta que explique si todos los animales usan el mismo método o usan diferentes, y qué método usa el sistema nervioso de cada animal. Obtener esa información está fuera del alcance de la investigación de esta tesis doctoral y la propia comunidad científica está lejos de obtener toda esa información. Así, puesto que hasta el momento solo existen propuestas sobre cómo las redes neuronales biológicas codifican la información la pregunta es no abordable.

Existen dos disciplinas en las que se ha generado interés por conocer cómo las redes neuronales almacenan información. Una de las disciplinas es la informática que busca crear y diseñar redes neuronales artificiales capaces de realizar tareas computacionales. En esta línea de investigación, cualquier red neuronal es relevante debido a que lo que importa es los problemas que son capaces de resolver. Así, las redes neurona-

les son dispositivos computacionales, por lo que su poder computacional y el proceso computacional que ellas realizan son investigados. La otra disciplina interesada es la neurociencia, que está interesada en la actividad de redes neuronales biológicas. La diferencia entre las líneas de investigación de cada una de las dos disciplinas proviene del problema a que se enfrenta cada una. En la primera, la red neuronal artificial es diseñada para que intente llevar a cabo tareas como: reconocimiento de patrones, regresiones o clasificaciones. Así, la informática se ocupa de diseñar redes neuronales e investigar su desempeño respecto a tareas computacionales. En la segunda, se conoce el desempeño que llevan a cabo las redes neuronales biológicas en tareas computacionales, pero se desconoce el mecanismo neuronal y la neurociencia trata de averiguarlo. Aunque, los objetivos de las dos disciplinas son diferentes, una interacción entre las dos líneas de investigación es muy positiva ya que puede producir sinergias que generan importantes avances en ambas líneas de investigación. Un ejemplo de esas sinergias son las redes de Hopfield que han sido un avance para ambas disciplinas [234].

El pensamiento consciente humano se presenta como un proceso simbólico en serie. Este hecho fue la inspiración para el trabajo de Turing en computación. El trabajo de Turing fue a la vez un elemento clave en el desarrollo de la psicología cognitiva. Sin embargo, la metáfora de la computadora no ayuda a entender como trabaja el cerebro porque la arquitectura de una máquina de Turing es muy diferente de la del cerebro humano. El cerebro humano tiene cientos de miles de millones de unidades de procesamiento que llevan a cabo un procesamiento masivo en paralelo. Allan Newell sugirió que un proceso computacional debería ser descrito mediante una descripción multinivel para poder alcanzar una completa comprensión de él [413]. Sin embargo, la propuesta de Newell está enfocada en su concepto de sistema de símbolos. David Marr propuso aplicar una descripción multinivel al sistema visual [341]. Concretamente, propuso tres niveles y desarrolló algoritmos concernientes a cada uno de los niveles. La idea de explicar el sistema nervioso usando una descripción multinivel ha sido también defendida por Mira y Delgado [394, 392].

En base a todo lo anterior, este capítulo se centra en repasar los métodos que se han propuesto para interpretar dinámicas cerebrales y presenta un mecanismo neuronal para interpretar dinámicas neuronales basado en algunos resultados de lenguajes formales producidos por la informática. Ese mecanismo identifica un estado interno o externo  $s$  con un lenguaje formal  $L$ . Independientemente, de si el mecanismo es usado o no en el cerebro de los seres humanos, el mecanismo puede ser usado para diseñar nuevas arquitecturas de redes neuronales artificiales. La organización del resto del capítulo es la siguiente: la segunda sección explica el problema sobre entender dinámicas neuronales desde el método de las descripciones multinivel. La tercera sección revisa los dos tipos de mecanismos que han sido propuestos hasta la fecha para interpretar la actividad neuronal. En la sección cuarta se hace un breve resumen de conceptos pertenecientes a lenguajes formales, redes neuronales y autómatas finitos deterministas. La sección quinta propone un nuevo mecanismo de categorización que podría ser responsable de

dinámicas cerebrales. Y la última sección tiene un resumen y una discusión breve del capítulo.

## 6.1. La descripción multinivel

Allan Newell propuso la idea de que los procesos computacionales que lleva a cabo un agente son descritos mediante una jerarquía de lenguajes [413]. La jerarquía se trataría de la jerarquía estándar de niveles ingeniería de computadores más un nivel adicional en el ápice, denominado nivel de conocimiento. La figura 6.1 muestra un diagrama de la jerarquía que propuso Newell.

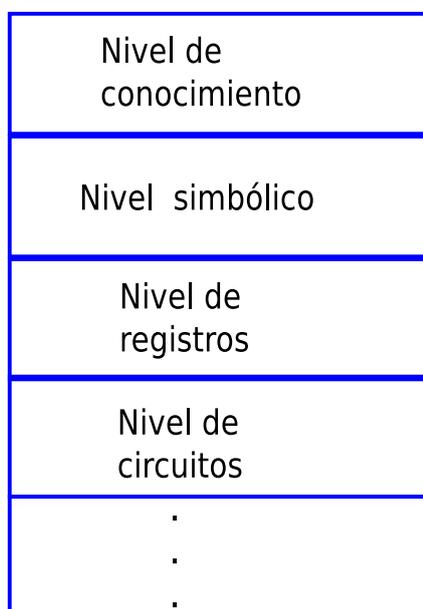


Figura 6.1: Niveles de la jerarquía de Newell

Newell explicaba que cada nivel de la jerarquía puede ser definido de dos maneras:

*“First, it can be defined autonomously, without reference to any other level to an amazing degree, programmers need not know logic circuits, logic designers need not know electrical circuits, managers can operate at the configuration level with no knowledge of programming, and so forth. Second, each level can be reduced to the level below. Each aspect of a level - medium, components, laws of composition and behavior - can be defined in terms of systems at the level next below. The architecture is the name we give to the register-transfer level system that defines a symbol (programming) level creating a machine language and making it run as described in the programmers manual for the machine. Neither of these two definitions of a level is the more fundamental. It is*

*essential that they both exist and agree.* ” [413]pag. 5.

El pensamiento consciente humano está en el nivel simbólico. El cómo el nivel simbólico emerge a partir del nivel inferior en las computadoras se conoce perfectamente ya que las computadoras son diseñadas por los seres humanos. Sin embargo, el cómo el nivel simbólico emerge del nivel inferior es una cuestión sin resolver. David Marr propuso que una jerarquía de tres niveles debería ser aplicada para comprender el sistema visual. La propuesta de Marr se podría aplicar a todo el sistema nervioso. Partiendo de ese hecho, José Mira Mira y Ana Delgado definieron una jerarquía con tres niveles principales: el nivel físico, el nivel simbólico y el nivel de conocimiento. [394, 392, 391]. Su propuesta defiende el uso de la jerarquía de niveles para entender la función del sistema nervioso. Una de las cuestiones que han discutido es cómo el nivel simbólico emerge a partir del nivel físico. En sus trabajos, argumentan que conocer todo sobre el proceso en el nivel físico en las redes neuronales no es suficiente para conocer lo que está siendo calculado en el nivel simbólico. Así, escriben lo siguiente:

*“Let us assume for one moment that we have a complete theory of the physical level. In other words, that we know everything about individual signals, circuits and local operators, similar to our knowledge about digital electronics and computer architecture. Would we know what the brain is calculating? Would we know the ‘program’ and the emerging cognitive processes? Of course, not.”* [394]pag. 224.

Por lo tanto, para moverse de un nivel a el siguiente se necesita un conjunto de reglas que permita la traducción. Un concepto importante que introducen en su propuesta Mira y Delgado es el concepto de ‘símbolo dinámico’ [394]. Mira y Delgado definen su concepto de ‘símbolo’ muy claramente dando una definición de símbolo que incluye a los sistemas neurofisiológicos.

*“ The symbols[in neurophysiological systems] are active and dynamic entities associated with specific patterns of spatio-temporal signals(electrical, chemical and electronics) that are presented repeatedly in a stable, independent and autonomous way and associated with specific references in the organism’s internal and external environment.”* [394]pag. 227.

Su propuesta expande explícitamente el concepto de símbolo empleado por Newell porque ellos desean aplicar una descripción multinivel a todos los tipos de sistemas computacionales, incluidos los sistemas conexionistas.

*“ we should remember that the concept of symbol proposed here is not the one proposed by Newell and Simon (1976) in their Physical Symbol System Hypothesis because in our proposal the symbols are evolutive, dynamic, connectionist and grounded in speci-*

*fic physiological mechanisms and they are not programmable in the conventional sense, but via dynamic adjustment processes of synaptic efficiency, controlled by the states of activity in the network. Per contra, symbols in conventional programmable computers are static, descriptive, and arbitrary semantics.” [394]pag. 228.*

Mira y Delgado también expusieron la situación problemática en la que se encuentra la comprensión de la función global del cerebro:

*“We know how to construct models at the physical level, in terms of mathematical relationships between measured physical quantities as function of time, but we still do not know how to construct formal models of symbols and relations between these symbols with the exception of those peripheral situations in which the neural networks are situated at end of the effector and/or receptor sides of the nervous system, with accessible points of physical measurement. It seems clear to us that we cannot use the conventional programming strategies as used in computers, because our symbols are embodied in the neural dynamics.” [394]pag. 229.*

Hasta hace pocos años, la investigación que se realizaba de la actividad neuronal de organismos biológicos era desarrollada fundamentalmente en el nivel físico, ya que registrar más de una neurona de una población era realmente complicado como se ha mencionado en la sección anterior. Solo en los últimos años la tecnología está permitiendo realizar experimentos en los que son registradas múltiples neuronas de una población al mismo tiempo. Sin embargo, las investigaciones teóricas en el campo de las redes neuronales han propuesto dos posibles mecanismos mediante los que traducir la actividad neuronal del nivel físico al nivel simbólico. La siguiente sección presenta brevemente esos dos mecanismos.

## 6.2. La actividad neuronal y la codificación de estados

Actualmente no se conoce cómo el cerebro codifica el pensamiento consciente. Sin embargo, diversos investigadores han propuesto mecanismos que hipotéticamente permitirían que poblaciones de neuronas pudieran codificar los estados simbólicos que componen el pensamiento consciente. Los mecanismos propuestos usan fundamentalmente uno de los dos siguientes métodos: vector de estado e itinerancia caótica. El concepto de vector de estado está detrás de diversos algoritmos de redes neuronales que incluyen métodos de memorias asociativas desarrollados por Kohonen, Anderson, Amari o Hopfield entre otros [271, 234] y las dinámicas de atractores desarrolladas por Amari, Hirsch, Hopfield o Amit entre otros [234, 18, 224, 20]. El concepto de vector

de estado propone representar un estado interno o externo mediante un vector cuyos componentes son definidos por el estado de las unidades de la red neuronal. El otro concepto que se ha formulado es el de itinerancia caótica [582, 583], propuesto en base a los comportamientos transitorios corticales. Algunos experimentos que se han llevado a cabo han mostrado actividad que contiene un conjunto de estados corticales que conmutan dinámicamente [264] y transiciones irregulares en el córtex visual de gatos [191]. Los proponentes de la itinerancia caótica argumentan en contra de la interpretación de las dinámicas relacionadas con el concepto de vector de estado (por ejemplo, el vector atractor) que no puede explicar los fenómenos transitorios mencionados [583]. Actualmente, no hay experimentos que resuelvan la disputa entre los dos mecanismos de interpretación. Por lo tanto, el cómo las dinámicas discretas de la actividad neuronal codifican las categorías perceptuales continúa siendo una cuestión abierta. Además, no hay razón para pensar que el vector de estado y la itinerancia caótica son los únicos mecanismos para producir representaciones de estados externos o internos en redes neuronales.

### 6.3. Lenguajes formales, redes neuronales y autómatas finitos deterministas

Como ya se ha mencionado en el capítulo 2, el trabajo de McCulloch y Pitts mostró que el sistema nervioso podría ser entendido como un sistema computacional [359]. El modelo de McCulloch y Pitts es equivalente a otro modelo computacional, el autómata finito determinista (AFD) [388]. Uno de los métodos para clasificar el poder computacional de un modelo computacional es compararlo con las clases de lenguajes definidas por la teoría de gramáticas generativas desarrollada por Noam Chomsky. El modelo de McCulloch y Pitts y el AFD pueden ser vistos como aceptadores de lenguajes, y así ser ubicados respecto a los lenguajes generados por gramáticas generativas. Una de las grandes ventajas de las redes neuronales es la capacidad para aprender tareas complejas relacionadas con clasificación o reconocimiento de patrones, pero también tiene desventajas. Una de ellas, es que el conocimiento aprendido por las redes neuronales es difícil de extraer y comprender. Esta desventaja puede ser vista en el problema del aprendizaje de lenguajes formales. Pero antes de explicar el problema recordemos algunas definiciones de los lenguajes formales. Un lenguaje formal  $L$  es un conjunto de cadenas de símbolos que podrían estar restringidas por un concreto conjunto de reglas. Una cadena,  $w$ , de un lenguaje es denominada *palabra*. El conjunto de símbolos que pueden estar en las palabras de un lenguaje se denomina alfabeto y se denotará por  $\Sigma$ . La longitud de una palabra es el número de símbolos que contiene, y se denota por  $|w|$ . Por lo tanto, siendo la palabra  $w = 101$  sobre el alfabeto  $\Sigma = \{0, 1\}$ , entonces  $|w| = 3$ . En los lenguajes formales se distingue entre dos tipos de lenguajes: los que

tienen palabras finitas y los que tienen palabras infinitas. Las gramáticas generativas de Chomsky permiten la definición de lenguajes de palabras finitas. El lenguaje con todas las palabras sobre un alfabeto  $\Sigma$  es denotado por  $\Sigma^*$ . Así, se puede decir que un lenguaje formal  $L$  sobre un alfabeto  $\Sigma$  es un subconjunto de  $\Sigma^*$ .

La IA y la ciencia cognitiva se han preocupado del problema del aprendizaje de lenguajes formales. Una de las técnicas para resolver este problema ha sido el uso de redes neuronales recurrentes. Algunos investigadores han propuesto que las redes neuronales recurrentes que reconocen lenguajes formales están simulando un AFD. Así, una red neuronal recurrente podría ser entendida generando un AFD a partir de la propia red neuronal. Este proceso se denomina extracción del AFD. El éxito de la extracción del AFD a partir de la red neuronal es cuestión abierta [272]. Sin embargo, un hecho sobre el que fijarse es que desde el modelo de McCulloch y Pitts hasta las redes neuronales dinámicas recurrentes todas pueden relacionarse con los lenguajes formales.

## 6.4. Un nuevo método de codificación neuronal: TLE

En esta sección se propone un nuevo mecanismo neuronal para la clasificación de señales de entrada. Las bases de este nuevo mecanismo son las siguientes:

1. La clasificación es un proceso fundamental de la cognición. Hay diferentes niveles de clasificación; y cuanto más abstractas son las categorías que puede crear un organismo, mayores son sus habilidades cognitivas. Olvidando las diferencias cualitativas del proceso de categorización en distintos animales, se trata de un proceso que poseen todos los animales, y que se usa en diversas funciones que van desde detectar peligros y sobrevivir [466] hasta llevar a cabo complejos procesos cognitivos [454].
2. Existen dinámicas discretas en el cerebro animal para la categorización. La investigación experimental ha mostrado que el disparo en las capas superficiales del córtex auditivo está organizado en un pequeño número de grupos neuronales que actúan como atractores, cuya respuesta puede predecir la discriminación de sonidos por parte del animal [38].
3. Una población de neuronas puede ser interpretada como un sistema dinámico. Las redes neuronales han sido usadas para modelar sistemas dinámicos [636] pero una red neuronal podría ser considerada un sistema dinámico también. Así, la continua entrada de señales desde su entorno que recibe un animal y las conexiones recurrentes pueden convertir una población neuronal en un sistema dinámico.
4. La secuencia temporal de estados producida por un sistema dinámico puede ser considerada una palabra. Una palabra es una secuencia de símbolos; por lo tanto,

si los estados de un sistema dinámico se consideran un alfabeto, entonces la secuencia de estados producida por un sistema dinámico puede ser visto como una palabra.

5. Los estados de consciencia están compuestos de una secuencia de señales de entrada con una duración finita. Diversas investigaciones han mostrado que hay un umbral mínimo de tiempo que es necesario superar para percibir un estímulo [249, 521].
6. Un lenguaje formal puede ser interpretado como una categoría cuyos objetos son palabras.
7. Las redes neuronales puede ser capaces de reconocer lenguajes formales [436]. Esto es relevante porque un lenguaje formal puede ser considerado como un elemento del nivel simbólico. Por lo tanto, esta conexión puede ser usada para ir del nivel físico a el nivel simbólico.

Dados los hechos (1) y (2) mencionados arriba, se debería encontrar un mecanismo capaz de categorizar si se desea explicar las dinámicas cerebrales. En base a los hechos (3) y (4), la actividad de las poblaciones de neuronas podría ser interpretada como un mecanismo que genera palabras pertenecientes a lenguaje formales. El hecho (5) implica que esas palabras deberían tener una longitud finita. Sin embargo, las palabras generadas por una red neuronal dinámica no serían interesantes si todas las palabras generadas pertenecen al mismo lenguaje formal, ya que sólo habría una categoría. Por lo tanto, tomando el hecho (6) y si se supone que el sistema dinámico genera palabras de diferentes lenguajes formales, una red neuronal puede asignar una categoría a una secuencia de señales de entrada asignándole un lenguaje formal. Esto es debido a que la trayectoria del sistema dinámico se puede interpretar como una palabra que pertenece a un lenguaje formal. Así, la actividad que denota un lenguaje formal específico se entendería como la asignación de una categoría. El hecho (7) muestra que las redes neuronales pueden aceptar o rechazar indicando si las palabras pertenecen o no a un lenguaje formal. Así, el mecanismo de la figura 6.2, denominado Trayectorias-Lenguajes-Estados(TLE), es propuesto como uno de los posibles mecanismos que puede producir dinámicas neuronales relacionadas con la clasificación.

El mecanismo consta de dos redes neuronales conectadas consecutivamente que pueden ser interpretadas en el nivel simbólico de la siguiente manera. La primera red transforma las señales de entrada en palabras mediante la generación de orbitas. La segunda red toma la palabra y reconoce a qué lenguaje formal pertenece. Así, mientras usualmente un AFD o una red neuronal son usados para reconocer palabras de un único lenguaje formal mediante la aceptación o rechazo de las palabras que se le presentan, aquí el objetivo es una red neuronal que reconozca el lenguaje formal al que pertenecen las palabras. Por lo tanto, la segunda red neuronal podría consistir en un conjunto

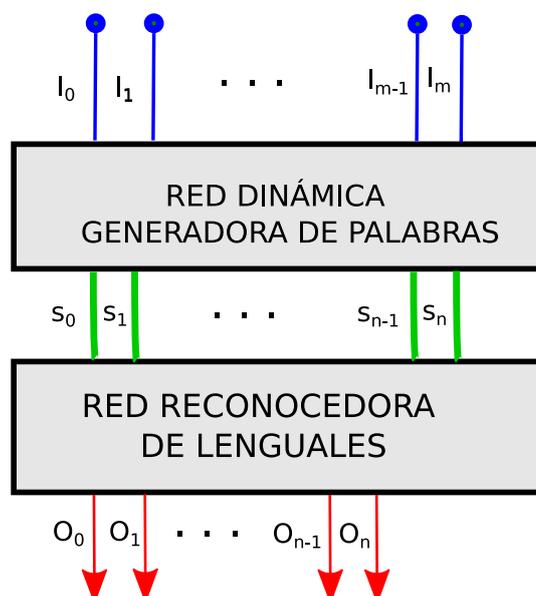


Figura 6.2: La figura muestra el esquema del mecanismo Trayectorias-Lenguajes-Estados propuesto para explicar dinámicas de redes neuronales relacionadas con la categorización.

de redes donde cada una está dedicada a reconocer un lenguaje formal y la salida de cada reconocedor conforma un estado vector que determina un lenguaje formal entre un conjunto de ellos. Así, una categoría es definida como un conjunto de trayectorias, pero cuando la categoría debe ser denotada se hace mediante un estado que es interpretado como el identificador de un lenguaje formal a nivel simbólico. Además, a la capacidad de categorización, el mecanismo propuesto puede servir como filtro de ruido. Un sensor puede obtener diversas variaciones de un mismo estado del entorno debido a la existencia de ruido o fluctuaciones del entorno. Así, puede ser necesario eliminar el ruido o las fluctuaciones. Si el número de lenguajes que pueden ser reconocidos es similar al tamaño del alfabeto entonces un lenguaje puede ser usado para representar un estado tras la eliminación de ruido en las señales de entrada.

Una de las cuestiones interesante es que el mecanismo TLE es escalable porque puede ser reaplicado en otros niveles. Por ejemplo, la secuencia temporal de lenguajes reconocidos a su vez conforma una cadena que puede ser interpretada también como una cadena de símbolos y por lo tanto una palabra de otro lenguaje formal. Así, se tendría un lenguaje formal cuyo alfabeto son los lenguajes formales del nivel anterior. Así, una jerarquía de lenguajes formales puede ser definida. Esa jerarquía representa que cada vez que se sube un nivel, las categorías asociadas con los lenguajes son más abstractas que en el nivel anterior.

## 6.5. Análisis de los resultados

Los datos experimentales actuales no permiten abordar las preguntas de ¿cómo representan estados los sistemas nerviosos de los animales? y ¿qué proceso evolutivo ha determinado la representación de información en las redes neuronales biológicas? Aun así, el conocer cómo el cerebro codifica los estados internos y externos para tomar decisiones y llevar a cabo exocomportamientos es fundamental para conocer los valores de fasa que poseen los seres humanos y otros animales. Pese a la falta de datos experimentales, las investigaciones teóricas han propuesto dos posibles mecanismos para explicar cómo las redes neuronales pueden codificar los estados del pensamiento humano: el vector de estado y la itinerancia caótica. Así, esos dos tipos de mecanismo son candidatos para explicar las dinámicas cerebrales. Ichiro Tsura considera que el modo en el que el cerebro codifica la información es mediante itinerancia caótica.

*“Furthermore, another misleading theory in conventional brain theory is the theory based on the description of non-stationary and transitory processes by a geometric attractor.”* [583] pag.2

Sin embargo, ya se ha mencionado que la investigación de Brice Bathellier y sus colaboradores [38] ha mostrado que las capas superficiales del córtex auditivo contienen un número pequeño de grupos de neuronas que actúan como atractores. Por lo que la naturaleza de los mecanismos neuronales que producen dinámicas cerebrales son una cuestión aún abierta.

Este capítulo ha presentado un nuevo mecanismo neuronal que podría subyacer en el funcionamiento del cerebro que se ha denominado trayectorias-lenguaje-estado. La idea expresada por este nuevo concepto es que un conjunto de trayectorias agrupadas sirve para definir una categoría. Este mecanismo puede ser traducido a nivel simbólico. Así, en el nivel simbólico las trayectorias son palabras de un lenguaje formal y el símbolo es formalizado como el lenguaje formal. Por lo tanto, un lenguaje formal llega a ser un símbolo para un estado interno o externo.

Los experimentos han demostrado dos hechos:

- Se ha encontrado actividad que contiene un conjunto de estados corticales de conmutación dinámica [264] y transiciones irregulares en el córtex visual [191].
- Disparo en las capas superficiales del córtex auditivo que está organizado en un número pequeño de grupos de neuronas actúan como atractores. [38].

Los dos hechos que se han citado, aunque desde la discusión entre el vector de estado y la itinerancia caótica por interpretar la actividad cerebral son aparentemente contradictorios, pueden ser reconciliados usando el mecanismo de TLE. Si se usa un

enfoque ascendente para analizar las dinámicas neuronales basado en el mecanismo TLE, se pueden explicar los resultados obtenidos en los experimentos de la siguiente manera. La continua generación de palabras por una red dinámica explicaría el cambio dinámico o las transiciones irregulares, y los atractores aparecerían en la red de los reconocedores de lenguajes. Por supuesto, esta explicación es solo una hipótesis, y se tendrían que llevar a cabo complejos experimentos para validarla, y es difícil saber si con la tecnología actual se pueden realizar esos experimentos. En el caso de que el mecanismo TLE existiera en el cerebro, los estados que representan el lenguaje formal serían los símbolos del pensamiento humano. El cómo esa actividad llega a ser consciente está probablemente relacionado con procesos físicos que aún no se entienden [316]; pero es una cuestión que queda fuera de los objetivos de esta tesis.

Hay una cuestión interesante que se debe notar sobre el mecanismo TLE, y es que este mecanismo puede incorporar los mecanismos de vector de estado e itinerancia caótica, ya que estos mecanismos pueden ser usados en las redes que forman la red del TLE. No parece razonable hablar de que un mecanismo neuronal es verdad y los otros no en su intento por explicar las dinámicas cerebrales, ya que es posible que el cerebro este usando más de un mecanismo para codificar información. El cerebro podría usar mecanismos diferentes a la vez si esto le permite optimizar su funcionamiento y operar de una manera más eficiente. Además, es posible que diferentes especies animales usen mecanismos neuronales diferentes para representar informaciones internas o del entorno. De cualquier modo, independientemente de si el método TLE existe en organismos biológicos, el nuevo mecanismo es interesante porque puede ser implementado en redes neuronales artificiales para tareas de clasificación o filtrado de ruido.

Finalmente, de la existencia de diferentes mecanismos de codificación para representar estados internos o externos se infiere que el modelo de red neuronal posee una gran versatilidad. Esta conclusión debe tenerse en cuenta a la hora de generalizar a otras especies el descubrimiento de un mecanismo neuronal de codificación en una especie animal concreta, ya que uno de los elementos fundamental de la evolución es la variabilidad.

## Capítulo 7

# La fasa y el sistema nervioso II: Robusted computacional

En el capítulo anterior se razonó que si el valor de fasa que posee un sistema queda conformado por procesos evolutivos, entonces los valores de los parámetros que determinan un valor de fasa son causados por procesos evolutivos. Así, se planteó que, dado que uno de los parámetros de la fasa es el modelo computacional, la existencia únicamente del modelo de red neuronal en los animales debe de poder ser explicado por procesos evolutivos. Así, tanto desde el punto de vista de la TGE, como desde la propia teoría de la evolución, se puede investigar qué proceso evolutivo ha conducido a que el modelo computacional sea el mismo en todos los organismos. Hasta la fecha, simplemente se ha asumido que el modelo de red neuronal es necesario para generar comportamiento en los organismos multicelulares complejos; pero la informática teórica muestra que existen muchos modelos computacionales alternativos. Los axones, los potenciales de acción, las dendritas y las sinapsis no son elementos que dirijan unívocamente a la aparición del modelo de red neuronal, ya que otros modelos computacionales podrían usar esos elementos. Además, las redes químicas pueden también procesar información [540, 226]. Dado que la variabilidad es una característica de la reproducción de organismos y un elemento clave en la evolución por selección natural [238], y existen múltiples modelos computacionales que no son el de red neuronal, se puede suponer que el modelo de red neuronal debería no ser el único en haber emergido en los animales. Por lo tanto, la principal pregunta que se va a abordar en este capítulo es: ¿por qué se encuentra solo el modelo de red neuronal en los clados de organismos multicelulares complejos? Sin embargo, por qué la red neuronal ha persistido y por qué no hay eumetazoos con otro tipo de modelo computacional están claros. Así, se puede preguntar legítimamente: (1) ¿la causa de que todos los eumetazoos tengan sistemas nerviosos neuronales es debido a un proceso de deriva aleatoria o de selección natural? (2) ¿Hay principios generales en la naturaleza que guíen la selección de un tipo de organismos frente a otros debido a las características de su mecanismo de toma de decisiones? Y (3) si es así, ¿cuáles

son esos principios? Responder a estas cuestiones permitiría obtener una mejor imagen de la evolución del sistema nervioso.

La respuesta a la pregunta (1) es que la existencia de sistemas nerviosos neuronales es causada por selección natural y no por deriva aleatoria. Una justificación inicial puede ser deducida del mismo concepto de deriva aleatoria. Normalmente, la deriva aleatoria tiene profundos efectos en la evolución en poblaciones pequeñas donde un pequeño grupo puede desaparecer por sucesos aleatorios que acaban con él. Sin embargo, la situación que se está abordando es completamente diferente, ya que sabemos, que los organismos multicelulares pueden surgir rápidamente de organismos unicelulares eucariotas [479], y los océanos eran un entorno homogéneo de seres unicelulares cuando los organismos multicelulares aparecieron [297, 89]. Así, la aparición de solo organismos multicelulares en un entorno espacial pequeño con pocos organismos de cada tipo que hubiera permitido un proceso de deriva aleatoria es improbable. Y sin esa situación, no pudo darse un proceso de deriva aleatoria con profundos efectos sobre la evolución de los organismos multicelulares que haya conducido a la persistencia únicamente del modelo de red neuronal. La evolución de los organismos unicelulares a los multicelulares debería haber ocurrido en múltiples lugares en los océanos. Así, la probabilidad de que la existencia del sistema nervioso con el modelo de red neuronal haya sido causada por la deriva aleatoria es demasiado baja como para no considerar que la causa sea alguna propiedad que da una mayor adaptabilidad biológica al organismo. Esa probabilidad es incluso menor si se considera la hipótesis de la poligénesis de las neuronas. La revolución genómica ha puesto seriamente en cuestión el origen único del sistema nervioso [408, 407]. Los análisis más recientes han sugerido que no existen genes-marcadores panneuronales y pansinápticos [406]. Además, se ha concluido que la neurotransmisión ha evolucionado a partir de dos orígenes diferentes [408]. Por otro lado, acetilcolina, serotonina, histamina, dopamina, octamina y el ácido gamma aminobutírico (GABA) fueron reclutados como transmisores en el sistema nervioso de los linajes de cnidarios y bilaterales, y que los ctenóforos evolucionaron la secreción de péptidos independientemente, sugiriendo que sistemas neuronales tempranos podrían ser peptidérgicos. Estos hechos sugieren que tanto la sinapsis eléctrica como la química han evolucionado más de una vez. Así, los nuevos datos que sugieren orígenes independientes y una evolución paralela de sistemas neuronales evitan que la deriva aleatoria pueda ser considerada una causa plausible de la persistencia del sistema nervioso. Asumiendo, la hipótesis de la poligénesis, la persistencia del sistema nervioso es un caso clásico de convergencia evolutiva, y, por lo tanto, debería haber una ventaja evolutiva que el modelo de red neuronal posee frente a otros tipos de modelos computacionales. Si la existencia de sistemas de procesamiento neuronales es causada por la selección natural, entonces la respuesta a la cuestión (2) debe ser, sí, existen principios, y esos principios deben ser encontrados para responder a la pregunta (3). Si la respuesta a (3) es encontrada, entonces se tendrá una explicación sobre por qué la deriva aleatoria no es la causa de la persistencia de sistemas neuronales de procesamiento en la naturaleza.

Por otro lado, si estuviéramos equivocados y la persistencia del modelo de red neuronal es causado por deriva aleatoria no se encontraría una ventaja evolutiva. Así, la cuestión de qué ventaja evolutiva posee el modelo de red neuronal, es la clave para responder la principal cuestión de este capítulo.

El punto de partida para investigar por qué la naturaleza selecciona el modelo de red neuronal como un mecanismo para hacer decisiones de comportamientos será la propuesta hecha por Rodolfo Llinás in 1987 [314]. Mientras él discutía la pregunta de cuál es la función de la mente, propuso que la respuesta a esa cuestión podría ser encontrada respondiendo a ¿para qué es bueno el cerebro? Su respuesta a esta última pregunta fue que el denominador común de todos los cerebros es la predicción, que es la última y más general de todas las funciones del cerebro. Él dedujo que la necesidad de predecir es causada por la habilidad de moverse porque sin predicción, moverse puede ser muy peligroso. Así, concluyó que moverse requiere predicción para evitar consecuencias negativas o lograr beneficios para el organismo. Debido a que han pasado muchos años desde que la propuesta fue hecha, se va a discutir la validez de la propuesta de Llinás. Además, se considerarán las características computacionales que todos los sistemas nerviosos comparten y cómo las características computacionales y de procesamiento de información conducen al proceso de selección natural en vez de enfocarse en la evolución concreta del sistema nervioso de un linaje.

Este capítulo está organizado de la siguiente manera: la sección 7.1 presenta una terminología que ayudará en la discusión de mecanismos de decisión que se encuentran lejos en el árbol evolutivo. A continuación, la sección 7.2 lleva a cabo una discusión de la propuesta de Llinás sobre la función del cerebro, contrastándola con descubrimientos relevantes hechos durante los últimos 30 años sobre diferentes organismos. La sección 7.3 da una descripción de las características que todos los cerebros neuronales tienen en común, y la sección 7.4 afronta la principal cuestión de este capítulo, es decir, ¿por qué el modelo de red neuronal es seleccionado frente a otros modelos computacionales en la naturaleza? Finalmente, la sección 7.5 analiza el resultado obtenido.

## 7.1. Comparando sistemas nerviosos

Los libros de texto de neurociencia dividen el sistema nervioso de un organismo en sistema nervioso central (SNC) y sistema nervioso periférico (SNP). La definición que se hace del SNC se hace mediante términos anatómicos. Por ejemplo:

“CNS[Central Nervous System] consists of the parts of the nervous system that are encased in bone: the brain and the spinal cord.”[340]p.171.

Este tipo de definiciones pierden su precisión cuando uno se aleja del subfilo verte-

brados [186], y las definiciones morfológicas no son útiles cuando se estudia o compara organismos alejados entre sí en el árbol evolutivo. Cuando se discute sobre el sistema nervioso de organismos que no pertenecen a los vertebrados se emplea una definición más general de SNC. La definición que se usa es que el SNC es la localización donde la concentración de neuronas es considerablemente mayor y su interrelación es más compleja que en otros lugares del organismo. Aunque esta segunda definición es más útil para estudiar o comparar los sistemas nerviosos de dos organismos alejados evolutivamente, todavía está basada en conceptos anatómicos que no permiten explícitamente la definición más generalizada. Esto puede llevar a problemas porque un grupo de neuronas interrelacionadas de un modo complejo puede categorizarse diferentemente dependiendo de su localización.

“In the CNS, such a group is called a nucleus and in the peripheral nervous system (PSN), a ganglion.” [79] p.5.

En algunos organismos un ganglio es considerado el cerebro del organismo, pero en los mamíferos, un ganglio es un grupo de neuronas localizadas fuera del SNC. Por lo tanto, el significado de la terminología depende del organismo que está siendo descrito o discutido, lo que es una limitación muy seria para discutir y comparar organismos muy distantes, hablando evolutivamente. La terminología para comparar y discutir un conjunto de organismos necesita ser tan general como grande sea la distancia entre los organismos en el árbol evolutivo. Por ejemplo, si se quiere comparar el sistema que genera comportamiento para una bacteria con el sistema que genera comportamiento en los mamíferos, se necesitarán conceptos funcionales, ya que las anatomías difieren completamente. Si se quieren comparar dos organismos que tienen sistema nervioso, pero están alejados en el árbol evolutivo, entonces no se pueden usar solo conceptos funcionales porque la terminología es demasiado general.

Delcomyn ha considerado un criterio funcional para definir un cerebro, afirmando que: “*brains process information, organize motor outputs, and make executive decisions*” [122] p.648. La definición funcional de Delcomyn ha sido usada para decidir si un ganglio de un organismo que no es un mamífero puede ser considerado un cerebro o no [95]. Sin embargo, esta definición de Delcomyn no es suficientemente precisa si la aplicamos a todos los sistemas físicos. Por ejemplo, se podría concluir que un robot tiene un cerebro porque tiene un sistema que cumple con la definición de Delcomyn aunque no tenga neuronas. Así, la definición de cerebro no debería ser solo desde un punto de vista funcional. Por lo tanto, es importantante no solo considerar qué información está procesando el sistema sino también cómo la procesa. Así, para abordar la principal cuestión que se ha formulado en este capítulo, se necesita distinguir un sistema nervioso que usa el modelo computacional de red neuronal del resto de sistemas con modelos computacionales diferentes. Basándose en lo anterior, se puede concluir que

es necesario una clasificación que permita diferentes niveles de precisión determinando cuándo y por qué dos sistemas nerviosos pertenecen a la misma clase para discutir la principal cuestión de este capítulo. Sin una clasificación de niveles, surgen problemas para discutir el sistema nervioso de organismos dado que la discusión concierne a organismos que viven en un mismo entorno ecológico, pero están alejados desde el punto de vista evolutivo. Así, una nueva clasificación para abordar la cuestión es presentada en la figura 7.1. Esta clasificación establece diferentes niveles. Cuanto mayor es el nivel, mayor es el número de clases en el cual un sistema nervioso puede ser dividido. Cada clase dentro de cada nivel es dividida en clases disjuntas en el siguiente nivel. El primer nivel de la clasificación solo tiene una clase que contiene todos los sistemas que existen desde el punto de vista de la teoría de la información. Los siguientes niveles son especificados mediante conceptos de la teoría de la información y computacionales. Si ninguno de los siguientes niveles que se ha mencionado es suficientemente poderoso para especificar la clase de sistemas que se busca, entonces se crean niveles superiores añadiendo características físicas para dividir las clases. En la neurociencia, la teoría de la información ha sido usada para analizar la actividad del sistema nervioso, pero la definición de los elementos del sistema nerviosos ha sido basada en las características anatómicas.

En el nivel 0 hay solo una clase, que contiene todos los sistemas que pueden ser considerados como un sistema nervioso desde el punto de vista funcional. La definición usada para determinar si un sistema es un sistema nervioso es la siguiente:

*Sistema nervioso: Es el sistema de un organismo que produce comportamientos mediante el cual el organismo interactúa con su entorno. Los elementos que registran información desde el entorno y los elementos que ejecutan el comportamiento no son parte del sistema nervioso.*

En el nivel 1, la clase sistema nervioso es dividido en tres clases: los sistemas de generación, los sistemas de transmisión, y sistemas de procesamiento.

- Sistema de generación: El sistema emite señales sin que posea un mecanismo de entradas de señales desde fuera del sistema y en una situación sin ruido del exterior del sistema.
- Sistema de transmisión: El sistema recibe señales de entrada enviadas por un emisor y transportada a otro sistema exactamente como fue recibida en una situación sin ruido.
- Sistema de procesamiento: El sistema recibe señales de entrada que son transportadas hasta la salida del sistema, pero al menos una señal es modificada al circular a través del sistema haciendo que la señal que emite sea diferente a la que recibe sin que haya ruido exterior.

Estas tres clases proporcionan la clasificación más general para mecanismos que generan comportamientos desde el punto de vista de la teoría de la información. Se debe notar que un sistema generador se convierte en un sistema de procesamiento cuando se le añade un sistema transmisor que permite la entrada de señales o al revés, ya que el sistema resultante produce unas señales de salida que son función de las de entrada, pero modificadas por la actividad interna. Por lo tanto, de las tres categorías, la categoría de sistemas de procesamiento contiene todos los posibles sistemas computacionales que tienen entrada de datos. Todos los sistemas nerviosos de los animales pertenecen a la clase de sistemas de procesamiento porque todos registran información del entorno y generan su comportamiento en base a esa información.

En esta investigación, interesa distinguir los sistemas nerviosos compuestos de neuronas de otros que usen otros modelos computacionales; Por lo tanto, necesitamos especificar la clasificación anterior. Esta clasificación puede ser adaptada a diferentes grados de especificación por generar subclases usando características computacionales. Una primera división de las clases es considerar sistemas distribuidos y centralizados. Un sistema distribuido es un sistema en el cuál la función es llevada por unidades independientes. Un sistema centralizado es un sistema en el cual la función es llevada a cabo por unidades acopladas fuertemente. Por ejemplo, las unidades de un sistema centralizado comportan una misma fuente de alimentación. Sin embargo, esta división no es suficiente porque la categoría de sistemas distribuidos contiene más sistemas que los que usan el modelo de red neuronal. Así, es necesario preguntarse que es una neurona. A excepción de las células de artrópodos y vertebrados, reconocer una neurona en un organismo es difícil porque ni los potenciales de acción ni la sinapsis especializada son prerequisites absolutos que las identifiquen [407]. Teniendo en cuenta lo anterior, el reconocer neuronas en el estudio de organismos fuera del reino de los animales es una cuestión problemática. Por ello, se debe establecer criterios para definir una neurona y así poder discutir la evolución y la selección de sistemas nerviosos con modelo de computación de red neuronal. Dada la necesidad de un criterio por los problemas que pueden surgir a través de los diferentes reinos, se propone usar un criterio computacional. Así, se propone definir una neurona con las siguientes dos características:

- Es una unidad con señales de entrada y salida.
- La función de computación que realiza es simple.

Usando estos criterios, se define un sistema neuronal como un sistema cuyas unidades son neuronas.

El significado de “simple” en la anterior definición es que cada estado de la unidad es un estado final, es decir, no posee estados intermedios que le permiten realizar a la unidad un proceso computacional complejo. Así, la categoría de sistemas de procesamiento distribuidos puede ser dividida usando el concepto de computación simple en sus unidades, produciéndose las siguientes categorías:

- Sistema de generación neuronal(SGN): Se trata de un sistema de generación compuesto de múltiples unidades independientes de procesamiento simple que están interconectadas y generan una señal de salida, pero sin ninguna entrada de señales.
- Sistema de transmisión neuronal(STN): Se trata de un sistema de transmisión compuesto de múltiples unidades independientes que están interconectadas.
- Sistema de procesamiento neuronal(SPN): Se trata de un Sistema de procesamiento compuesto de múltiples unidades independientes de procesamiento simple que están interconectadas.

Debe notarse que, desde el punto de vista de la teoría de la información, las redes neuronales pueden tener una función completamente diferente con diferentes configuraciones. En el modelo computacional de red neuronal, cada entrada de una unidad es modificada por un peso, que usualmente multiplica el valor de la señal de entrada. Si el peso no modifica la entrada, la red neuronal es un sistema de transmisión. Dividir el modelo de red neuronal en tres tipos diferentes permite distinguir con claridad cuando la función llevada a cabo por una red neuronal es completamente diferente de otra. Esto es crucial para distinguir los sistemas nerviosos de organismos multicelulares que son evolutivamente distintos.

La clase de los sistemas de procesamiento neuronal puede además ser refinada mediante su división en dos tipos: los multisistemas de procesamiento neuronal y los monosistemas de procesamiento neuronal, referenciando la existencia o no de redes independientes con poder computacional completo, respectivamente. Además, otros refinamientos se pueden realizar usando características físicas. En la investigación de este capítulo solo se necesita hasta el tercer nivel.

## 7.2. Sistemas nerviosos, predicción y la selección natural

Rodolfo Llinás ha propuesto que la función común de todos los cerebros es la predicción, y que la necesidad de predicción es causada por la habilidad de moverse [314, 315]. En la propuesta ha definido la predicción de la siguiente manera:

“ It is the ability to anticipate on the basis of incoming sensory stimuli the ultimate outcome of a given change in the external surroundings and to respond to such change with goal-directed actions that ensures survival.” [314] p.340

NIVEL	CLASES		
0	SISTEMA NERVIOSO		
1	SISTEMA DE GENERACION	SISTEMA DE PROCESAMIENTO	SISTEMA DE TRANSMISIÓN
2	SISTEMAS LOCALES DE GENERACIÓN	SISTEMAS LOCALES DE PROCESAMIENTO	SISTEMAS LOCALES DE TRANSMISIÓN
	SISTEMAS DISTRIBUIDOS DE GENERACION	SISTEMAS DISTRIBUIDOS DE PROCESAMIENTO	SISTEMAS DISTRIBUIDOS DE TRANSMISIÓN
3	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid red; width: 20px; height: 20px; display: inline-block;"></div> <div style="border: 1px solid red; width: 20px; height: 20px; display: inline-block;"></div> <div style="font-size: 10px;">...</div> <div style="border: 1px solid red; width: 20px; height: 20px; display: inline-block;"></div> </div> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid red; width: 20px; height: 20px; display: inline-block;"></div> <div style="border: 1px solid red; width: 20px; height: 20px; display: inline-block; text-align: center;">SSGN</div> <div style="font-size: 10px;">...</div> <div style="border: 1px solid red; width: 20px; height: 20px; display: inline-block;"></div> </div>	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid green; width: 20px; height: 20px; display: inline-block;"></div> <div style="border: 1px solid green; width: 20px; height: 20px; display: inline-block;"></div> <div style="font-size: 10px;">...</div> <div style="border: 1px solid green; width: 20px; height: 20px; display: inline-block;"></div> </div> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid green; width: 20px; height: 20px; display: inline-block;"></div> <div style="border: 1px solid green; width: 20px; height: 20px; display: inline-block; text-align: center;">SSPN</div> <div style="font-size: 10px;">...</div> <div style="border: 1px solid green; width: 20px; height: 20px; display: inline-block;"></div> </div>	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid blue; width: 20px; height: 20px; display: inline-block;"></div> <div style="border: 1px solid blue; width: 20px; height: 20px; display: inline-block;"></div> <div style="font-size: 10px;">...</div> <div style="border: 1px solid blue; width: 20px; height: 20px; display: inline-block;"></div> </div> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid blue; width: 20px; height: 20px; display: inline-block;"></div> <div style="border: 1px solid blue; width: 20px; height: 20px; display: inline-block; text-align: center;">SSTN</div> <div style="font-size: 10px;">...</div> <div style="border: 1px solid blue; width: 20px; height: 20px; display: inline-block;"></div> </div>
.	.		.
.			.
.			.

Figura 7.1: Esta tabla muestra los diferentes niveles y clases en las cuales el sistema nervioso puede ser clasificado. Cada clase del nivel  $n$  es dividida en clases disjuntas en el nivel  $n + 1$ . SSGN= Sistemas de generación neuronal; SSPN = Sistemas de procesamiento neuronal; SSTN = Sistemas de transmisor neuronal.

Una cosa que debe de notarse sobre esta definición es que la anticipación del resultado último de un cambio puede corresponder únicamente con la generación de una respuesta al cambio sin una representación explícita del futuro. A pesar de que la definición de una predicción es descrita en el diccionario de Cambridge como “a statement of what you think will happen in the future”, este significado hace referencia a la predicción que hacemos los seres humanos. Por otro lado, la definición dada por Llinás es más general porque no requiere una representación explícita del futuro. Esto es debido a que ya existe una representación implícita, supervivencia, asociada con la respuesta de los organismos biológicos. Esa asociación es creada por el proceso de selección natural. Nótese que la relación entre predicción y aprendizaje es un modo de ver claramente que el comportamiento de un organismo sin una representación explícita del futuro ha sido asociado con la predicción de supervivencia. El aprendizaje está relacionado con la predicción porque un organismo usa el aprendizaje para modificar su respuesta al enfrentarse a situaciones nuevas del entorno y sobrevivir. Por lo tanto, la respuesta de un organismo es debida a una predicción, aunque no tenga una representación explícita del futuro porque el proceso de selección natural garantiza que la respuesta proporciona al organismo una alta probabilidad de sobrevivir.

El artículo de Llinás afirmaba que la necesidad de predicción ha sido causada por las habilidades relacionadas con el movimiento y las grandes diferencias aparecen cuando uno se mueve en una dirección u otra. Llinás ha propuesto que el ciclo de vida del tunicado es una prueba que apoya la hipótesis de la libertad de movimiento. Los tunicados primitivos y algunas formas actuales tiene dos etapas muy diferentes durante su ciclo de vida: larva y la forma adulta. Las larvas son nadadores libres y poseen un ganglio semejante a un cerebro que recibe información de un estatocisto, un primitivo ojo y una primitiva médula espinal. La forma adulta es sésil; sus dos funciones principales son alimentarse filtrando agua y reproducirse. Las larvas nadan buscando un sustrato apto; cuando lo encuentran proceden a implantar sus cabezas en esa localización y cambiar a su forma adulta. Durante su metamorfosis a la forma adulta, la mayoría de su cerebro es absorbido. Llinás considera que eso es una prueba biológica de la relación entre movimiento y cerebro. Llinás afirma:

“It is that brains are needed only in those multicellular beings that move actively.”  
[314] pag.341

Esa afirmación tiene además un elemento implícito porque la propuesta resalta que la función del cerebro es la predicción. Así, la hipótesis puede ser sintetizada de la siguiente manera:

*Los cerebros se necesitan para predecir solo en aquellos organismos multicelulares que se mueven activamente.*

Pero hay algo más que menciona la propuesta:

“[multicellular beings that move actively] must interact with the environment at speeds that are sufficiently fast to require a rapidly acting organ specialized for the forecasting of external changes with sufficient anticipation (prediction) to afford predictive behaviour.” [314]p.343

El párrafo anterior es importante porque al considerarse que cada nicho implica diferentes necesidades de predecir y reaccionar conlleva que diferentes nichos requieren diferentes cerebros. Eso lleva a dos implicaciones: Una, cuanto más rápido es el cambio al que tiene que responder el organismo, más rápido se tiene que predecir, y dos, cuanto mayor es la complejidad entre las relaciones de los cambios del entorno y la respuesta que tiene que generarse, mayor es la complejidad de la computación requerida para hacer dicha predicción. Esto debe ser incluido en la síntesis. Así,

*Los cerebros se necesitan para predecir solo en aquellos organismos multicelulares que se mueven activamente. Los mecanismos y características del cerebro de un organismo para comunicar e integrar señales están directamente relacionados con la complejidad de la relación entre los cambios del entorno y las respuestas que debe dar el organismo, así como a la velocidad de dichos cambios.*

La afirmación anterior será denominada la hipótesis de la libertad de movimiento. Un lector se podría preguntar por qué se ha indicado “la velocidad de los cambios” en vez de “la velocidad de los movimientos del organismo”. La primera razón es porque hay dos opciones. (i) La velocidad de los cambios es causada por la velocidad del organismo. Por lo tanto, la velocidad de los cambios es causada directamente por la velocidad de los movimientos. (ii) La velocidad de los cambios es causada por la velocidad de la amenaza o la presa. En este caso, la velocidad de movimientos del organismo debe ser suficientemente rápida para evitar la amenaza o capturar a la presa porque si eso no ocurre el organismo no es seleccionado. Por lo tanto, la velocidad de movimientos del organismo está directamente relacionada con la velocidad de los cambios en el entorno en ambos casos.

A pesar de que la hipótesis de la libertad de movimiento aparenta ser sencilla, pueden aparecer ambigüedades sobre su significado o dudas sobre la propia veracidad de ella. Las cuestiones que pueden surgir sobre el enunciado de la hipótesis son las siguientes: ¿Qué es un “cerebro”?, ¿qué significa “se necesitan”?, ¿qué tipo de proceso lleva a cabo las acciones de anticipación y respuesta con referencia a “predecir”?, ¿qué significa “moverse activamente”?, ¿son las medusas un contraejemplo?, ¿es el uso de señales eléctricas por las plantas un contraejemplo?, ¿es la venus atrapamoscas un

contraejemplo?, ¿es una esponja un contraejemplo?, ¿es el *Trichoplax* un contraejemplo?, ¿un sistema nervioso es una condición suficiente y necesaria para predecir?, ¿es la capacidad de moverse una condición necesaria y suficiente para un sistema de procesamiento neuronal? y ¿puede la hipótesis de la libertad de movimiento explicar por qué el sistema de procesamiento neuronal es seleccionado frente a otros tipos de sistemas de procesamiento por la naturaleza? Ante esta lista de cuestiones, primeramente, se va a abordar las cuestiones relacionadas con las ambigüedades para después pasar a discutir el resto de cuestiones.

Primero, se comienza abordando qué criterio debe ser usado para determinar el significado de “cerebro” en el enunciado de la hipótesis. El cerebro está referenciando una alta concentración de neuronas al cargo de la función de generar el comportamiento mediante el registro de señales del entorno. Sin embargo, la definición de una alta concentración es ambigua. No parece una característica fundamental porque hay ejemplos de sistemas nerviosos en moluscos con una baja concentración [407]. Se ha demostrado que la alta concentración de neuronas es una consecuencia de la optimización para una anatomía concreta [97]. Por lo tanto, la alta concentración no es relevante para definir un cerebro. Además, tampoco la definición de neurona es precisa desde el punto de vista biológico porque hay muchas variaciones [407]. Estas ambigüedades pueden ser resueltas empleando la clasificación propuesta en la sección 7.1. El uso del término cerebro en el enunciado de la hipótesis implica dos cuestiones que deben ser tomadas en cuenta. Una es que el comportamiento es generado de acuerdo a unas señales de entrada. La segunda es cómo el sistema procesa información. El término cerebro puede ser precisado con la clase de sistemas de procesamiento neuronal (SSPN) como fue definida en la sección 7.1. Además, esto es compatible con la intención original de referenciar el SNC. Por lo tanto, se puede clarificar la hipótesis de la libertad de movimiento formulándola de la siguiente manera:

*Los sistemas de procesamiento neuronal se necesitan para predecir solo en organismos multicelulares que se mueven activamente. Los mecanismos y características del sistema de procesamiento neuronal de un organismo para comunicar e integrar señales están directamente relacionados con la complejidad de la relación entre los cambios del entorno y las respuestas que debe dar el organismo, así como a la velocidad de dichos cambios.*

En relación a la segunda cuestión, ¿qué significa “son necesitados”? Alguien podría pensar que la expresión “son necesitados” implica que lo que necesita un organismo es conocido por la naturaleza. Sin embargo, la naturaleza es como un jugador que lanza los dados que no conoce el resultado que va a salir, pero tampoco sabe que resultado es el requerido para ganar. Después de que el jugador ha lanzado el dado, el crupier le informará al jugador si ha ganado el juego o ha perdido; pero el jugador desconoce

con qué número será ganador. Por lo tanto, cuando se afirma que un organismo con la habilidad de moverse tiene la necesidad de predecir, sería un error interpretar que la naturaleza sabe de alguna manera que los organismos con la habilidad de moverse necesitan un cerebro. La afirmación de que un organismo con la habilidad de moverse tiene la necesidad de predecir significa que la adecuación biológica de una población de organismos con la habilidad de moverse sin un mecanismo de predicción en ese entorno es cero. Por lo tanto, los organismos para ser capaces de sobrevivir tienen que tener un mecanismo de predicción. Esto es, la necesidad de algo significa que sin ello el valor de adecuación biológica para el organismo es cero. Tomando en cuenta esta clarificación, la hipótesis de la libertad de movimiento puede ser reformulada mediante la sustitución de “se necesitan” por “son seleccionados”.

*Los sistemas de procesamiento neuronal son seleccionados para predecir solo en organismos multicelulares que se mueven activamente. Los mecanismos y características del sistema de procesamiento neuronal de un organismo para comunicar e integrar señales están directamente relacionados con la complejidad de la relación entre los cambios del entorno y las respuestas que debe dar el organismo, así como a la velocidad de dichos cambios.*

La siguiente cuestión con la que continuar es: ¿Qué tipo de proceso lleva las acciones de anticipar y responder referenciada por “predecir”? A pesar de que la definición de predecir no menciona cómo anticipar y responder son llevados a cabo, la afirmación de que la predicción es la principal función del cerebro, lo que es un sistema físico, previene la consideración de cualquier tipo de dualismo. Por lo tanto, predecir debería ser entendido como procesamiento de información del entorno para generar movimientos que permitan sobrevivir. Además, el procesamiento de información encaja con los dos significados de predicción, lo que incluye metas explícitas e implícitas. Dada esta nueva clarificación, una formulación menos ambigua de la hipótesis de la libertad de movimiento es la siguiente:

*Los sistemas de procesamiento neuronal son seleccionados para procesar información del entorno y generar movimientos para sobrevivir solo en organismos que se mueven activamente. Los mecanismos y características del sistema de procesamiento neuronal de un organismo para comunicar e integrar señales están directamente relacionados con la complejidad de la relación entre los cambios del entorno y las respuestas que debe dar el organismo, así como a la velocidad de dichos cambios.*

La siguiente ambigüedad de la que encargarse es: ¿Qué significa “moverse activamente”? La expresión se refiere a que el organismo está usando su propia energía. Hay

que notar que no todos los organismos se mueven activamente. Por ejemplo, las semillas de los dientes de león son contenidas dentro de aquenios que hacen que puedan ser transportadas largas distancias por el viento. Además, aunque no se ha mencionado explícitamente en el enunciado, la propuesta es sobre organismos que se encuentran en entornos heterogéneos, donde direcciones diferentes llevan a consecuencias diferentes, porque los entornos de la tierra donde los sistemas de procesamiento neuronal han sido seleccionados tienen estas características. Tomando todo esto, la hipótesis puede ser de nuevo reformulada del siguiente modo:

*En un entorno multidimensional heterogéneo, los sistemas de procesamiento neuronal son seleccionados para procesar información del entorno y generar movimientos para sobrevivir solo en organismos que usan energía interna para producir esos movimientos. Los mecanismos y características del sistema de procesamiento neuronal de un organismo para comunicar e integrar señales están directamente relacionados con la complejidad de la relación entre los cambios del entorno y las respuestas que debe dar el organismo, así como a la velocidad de dichos cambios.*

Ahora se abordará la última cuestión de ambigüedades: ¿Qué características de los sistemas de procesamiento neuronal están envueltas en la afirmación? Las características consideradas son todas aquellas que definen neuronas en el sistema nervioso. Sin embargo, a pesar de que se han propuesto diversas definiciones de qué es una neurona, lo cierto es que hay variación dependiendo de las características del animal. Por ejemplo, hay algunos metazoos inferiores y bilaterales cuyas neuronas carecen de sinapsis [407]. Además, todas las propuestas sobre la evolución del sistema nervioso consideran que se han ido añadiendo características y aumentado el nivel de complejidad desde los protosistemas nerviosos que permitieron comunicación e integración de señales [443, 442, 444, 198, 326, 423, 260]. Por lo tanto, las características no solo referencian la complejidad de las neuronas de los mamíferos sino también cualquier variación que haya existido o pueda existir en la evolución de los sistemas nerviosos. La definición computacional de sistema de procesamiento neuronal referencia solo unidades independientes que se comunican entre ellas e integran señales de otras para emitir sus propias señales. Así, todas las variaciones que han surgido durante la evolución del sistema nervioso cumplen la definición. Una vez aclarada la cuestión, la formulación de la hipótesis es la siguiente:

*Hipótesis de la libertad de movimiento: En un entorno multidimensional heterogéneo, los sistemas de procesamiento neuronal son seleccionados para procesar información del entorno y generar movimientos para sobrevivir solo en organismos que usan energía interna para producir esos movimientos. Los mecanismos y características del sistema de procesamiento neuronal de un organismo para comunicar e integrar señales están direc-*

*tamente relacionados con la complejidad de la relación entre los cambios del entorno y las respuestas que debe dar el organismo, así como a la velocidad de dichos cambios.*

Una cuestión a tener en cuenta es que la hipótesis de la libertad de movimiento no implica que un organismo multicelular con un SPN será seleccionado; sino que si no lo tiene no será seleccionado.

### **7.2.1. Revisando comportamientos, sistemas nerviosos y organismos**

Ahora que se ha obtenido una formulación sin ambigüedades de la hipótesis de la libertad de movimiento, se está en disposición de abordar las cuestiones que desafían la hipótesis. Sin embargo, una discusión de esas cuestiones implica un análisis biológico en profundidad de ellas. En la sección 2.4.2 se han revisado ya descubrimientos sobre plantas, en la 2.4.3 sobre hongos y en la 2.4.4 sobre microorganismos; pero son necesarios más datos. Así, a continuación, se va a revisar otros descubrimientos también relevantes para la posterior discusión de las cuestiones que se han planteado sobre la hipótesis de la libertad de movimiento.

#### **7.2.1.1. Optimización de redes**

Durante las últimas décadas, se han obtenido importantes resultados en la comprensión de la organización del sistema nerviosos aplicando métodos de optimización de redes [88, 290]. La minimización de conexiones disminuye la cantidad de tiempo y energía que se necesita para transmitir información. A partir de ese hecho se ha logrado explicar la organización del sistema nervioso y por qué el cerebro está en la cabeza de los vertebrados e invertebrados [97]. Así, las investigaciones realizadas han dado explicación de la forma y localización de diversos elementos del sistema nervioso.

#### **7.2.1.2. Las medusas, predicción y comportamiento**

El estudio de las medusas es importante porque es posiblemente el animal multicelular más antiguo. Se sabe que aparecieron hace más de 500 millones años [94]. Las medusas son animales marinos que nadan libremente. Las tres clases de medusas (Cubozoa, Hydrozoa and Scyphozoa) tienen sistema nervioso, pero ninguno presenta cefalización. Algunas medusas tienen una notable capacidad para detectar la dirección de las corrientes del océano y nadar fuertemente en contra de ellas. Las medusas son importantes en el análisis de la hipótesis de la libertad de movimiento porque se ha considerado que no poseen un SNC sino solo una red nerviosa. Así, alguien podría argumentar que las medusas se mueven activamente aun sin tener un cerebro. Sin embargo, esa visión

tradicional del sistema nervioso de la medusa se ha puesto en duda seriamente, así como la definición de cerebro [504, 175, 327, 231].

Diez años atrás se hizo un sorprendente descubrimiento sobre las medusas. Entre las medusas, se encuentra la medusa de caja. Las medusas de caja son vigorosas nadadoras y predadores activos [503]. Las medusas de caja difieren de otras medusas porque tienen un sistema sensorial elaborado y demuestran un comportamiento activo de caza. Las medusas de caja poseen un sorprendente sistema visual con una matriz interactiva de 24 ojos que son de cuatro tipos distintos. Los experimentos han mostrado que las lentes de las medusas de caja poseen un gradiente de índice refractario ajustado de manera muy fina que produce imágenes que están muy cerca de estar libres de aberraciones [427]. Sin embargo, los experimentos también han mostrado que la posición de la retina no coincide con la de la imagen nítida, llevando a campos receptivos complejos y amplios en los fotorreceptores individuales [427, 611]. Estos descubrimientos implican que la arquitectura de los ojos es útil sirviendo a una única tarea visual. Así, se trata de un ejemplo que muestra que la visión no es un sentido de propósito general.

### 7.2.1.3. Las esponjas, predicción y comportamiento

Durante mucho tiempo, se ha considerado que las esponjas no poseían un sistema nervioso. Sin embargo, esta visión ha sido reconsiderada durante los últimos años [304, 423, 481]. El principal comportamiento de las esponjas es filtrar agua y prevenir la entrada de partículas no deseadas. La esponja para sobrevivir necesita expulsar material no comestible que podría obstruir sus cavidades. Otro comportamiento es el control del diámetro del canal para variar la cantidad de agua procesada, reduciendo la cantidad de agua filtrada si el alimento está limitado. Además, tienen comportamientos de locomoción y contracción [62, 424], aunque son muy lentos y se necesitan técnicas de *time-lapse* para medirlos. Se ha descubierto que las esponjas tienen células sensoriales, vías de comunicación, y moléculas señalizadoras para generar esos comportamientos [304, 423, 481]. Las esponjas vítreas usan señales eléctricas [305], y el resto de esponjas parece que usan aminoácidos como transmisores para la coordinación de los comportamientos, aunque actualmente existen pocas herramientas que se puedan usar para investigar en profundidad esa cuestión [304]. Posiblemente, en los próximos años las investigaciones arrojarán más datos sobre el sistema nervioso de las esponjas.

### 7.2.1.4. El *Trichoplax*

El *Trichoplax* es un pequeño metazoo que produce su locomoción mediante deslizamiento ciliar. A pesar de que este organismo ha sido estudiado durante más de 40 años en laboratorios, se conoce poco sobre su comportamiento. Se sabe que solo tiene seis tipos de células y que carece de sinapsis, pero puede coordinar una secuencia de comportamientos culminando en una digestión externa de algas. El *Trichoplax*

detiene sus movimientos y deslizamiento ciliar en presencia de comida [534]. Ese comportamiento coordinado de células mientras se alimenta implica la existencia de señales intracelulares. Dos tipos de células en el *Trichoplax* son candidatas para controlar su comportamiento: las glandulares y las fibrosas [536]. Las células glandulares contienen neuropéptidos y podrían funcionar de una manera paracrina. Las células fibrosas tienen uniones que podrían permitir la conducción eléctrica y además tienen uniones con otros tipos de células; pero se desconoce si verdaderamente son eléctricamente excitables .

### 7.2.2. Confrontando la hipótesis de la libertad de movimiento

Tras la anterior revisión, ya se puede abordar las cuestiones que están pendientes desde el principio de la sección y que ponen en cuestión la veracidad de la hipótesis de la libertad de movimientos. Comencemos con la siguiente cuestión: ¿Son las medusas un contraejemplo de la hipótesis de la libertad de movimiento? Los estudios han mostrado que hay diferentes grados de compresión de la red neuronal y de condensación neuronal con estructuras semejantes a ganglios de las medusas. Así, la comunidad científica considera que el sistema nervioso de las medusas posee un SNC. Esto responde negativamente a la pregunta de si las medusas son un contraejemplo de la hipótesis de la libertad de movimiento. Además de los argumentos expuestos en la bibliografía, un punto que debe de ser tenido en cuenta es la optimización del sistema nervioso. Las altas densidades de neuronas encontradas en los SNC clásicos son debidas a la selección natural para obtener una optimización. Por lo tanto, la morfología del sistema locomotor predetermina la organización optima del sistema nervioso. Si el SNC de la medusa no está concentrado es porque no produce una ventaja evolutiva o incluso produce una desventaja; sin embargo, en vertebrados e invertebrados, sus morfologías conducen a la organización del sistema nervioso en concentraciones porque optimiza distintos parámetros. Adicionalmente, el sistema nervioso de la medusa puede ser clasificado como un SPN porque posee una entrada de señales y no sólo las transporta, sino que las modifica para producir el comportamiento de salida.

Otra cuestión es si el uso de señales eléctricas por plantas es un contraejemplo. Se ha mencionado que la bibliografía actual en botánica considera que las plantas tienen un sistema nervioso simple dentro de las células del floema [595, 162], el cuál es responsable de la comunicación a larga distancia [216, 635]. El papel del sistema nervioso en las plantas es enviar información [635]. Así, el sistema nervioso de las plantas no se trata de un sistema de procesamiento sino un sistema de transmisión. Su estructura es la de un sistema de transmisión neuronal porque los modelos de su sistema nervioso consisten en una matriz de células excitables [552]. Puesto que la hipótesis de la libertad de movimiento es sobre un SNP, no hay conflicto entre ello y la existencia de un sistema de transmisión en las plantas. Respecto a la hipótesis de la “raíz-cerebro”, estaría en conformidad con la hipótesis de la libertad de movimiento. Esto es debido a que si las células de la zona de transición del ápex de la raíz procesan información para tomar

decisiones de hacia dónde moverse se trataría de un caso que entra en los que establece la hipótesis de la libertad de movimiento.

La siguiente cuestión para abordar es: ¿es la venus atrapamoscas un contraejemplo? Esta es una cuestión diferente del caso del floema en las plantas. En la revisión sobre la venus atrapamoscas se vio que envía señales eléctricas para cerrar la trampa y capturar insectos. La venus atrapamoscas tiene células sensitivas que traducen los estímulos mecánicos en señales eléctricas, que son recibidas por células receptoras. La célula receptora procesa primero la señal usando el electrotono para predecir si el estímulo es causado por un insecto o no. Si la célula receptora produce un potencial de acción, este será transmitido para cerrar la trampa. Debe notarse que no todos los estímulos mecánicos causan que la trampa se cierre porque la venus atrapamoscas está procesando las señales que está recibiendo. Por lo tanto, no se trata de un sistema de transmisión de señales sino de un sistema de procesamiento. Así, desde el punto de vista del procesamiento de información se debe considerar que la venus atrapamoscas tiene un sistema de procesamiento neuronal porque procesa señales y sus unidades están interconectadas. En base a lo anterior, se concluye que la venus atrapamoscas cumple con la hipótesis de la libertad de movimiento. Algún lector podría indicar que su sistema de procesamiento es demasiado simple, pero esto no es un argumento válido en contra de la hipótesis porque su enunciado afirma que las características de un SPN de un organismo están relacionadas directamente con la velocidad de sus movimientos y la complejidad entre los cambios del entorno y las respuestas del organismo. Si la complejidad entre los cambios y las respuestas es pequeña, el grado de complejidad del SPN será bajo también. Las venus atrapamoscas podrían haber evolucionado para desarrollar un sistema visual para reconocer presas considerando solo la variabilidad reproductiva, pero los entornos donde existen, esa forma no es viable energéticamente. Si un organismo gasta más energía de la que puede obtener de su entorno, la selección natural no lo va a mantener. Se sabe que el sistema nervioso es muy costoso desde el punto de vista energético [228] y que las señales eléctricas requieren grandes cantidades de energía [19]. En el caso de los eumetazoos el balance es positivo. Sin embargo, una venus atrapamoscas con un sistema visual para capturar insectos no poseería un balance positivo. La naturaleza ha creado un sistema nervioso para las venus atrapamoscas optimizado de tal manera que toma decisiones y mantiene un balance energético positivo. No obstante, la venus atrapamoscas es un caso relevante para la hipótesis de la libertad de movimiento, ya que muestra que hay diferentes grados complejidad en el procesamiento y predicción. De hecho, ese es un hecho que en el campo de las matemáticas ya se conoce. Se sabe que los algoritmos para procesar señales multidimensionales son mucho más complejos que los que se necesitan para procesar una señal unidimensional. Esto se debe a varios hechos matemáticos tales como que las señales multidimensionales deben ser modeladas con ecuaciones diferenciales parciales, y que el teorema fundamental del álgebra no se cumple en ellas. Por lo tanto, puede ser formulado el siguiente principio evolutivo :

*Principio de procesamiento de señales multidimensionales: Se seleccionan sistemas de alto poder computacional para aquellos organismos cuyos nichos requieren reaccionar a señales multidimensionales para estar adaptados biológicamente.*

Otra duda que se puede plantear es si una esponja es un contraejemplo de la hipótesis de la libertad de movimiento. Como se ha mencionado en la revisión del apartado anterior, la esponja necesita procesar información del entorno para generar las acciones que la permitan sobrevivir. La señalización de una esponja es lenta en comparación con la señalización eléctrica de los sistemas nerviosos de los vertebrados, pero es sin duda efectiva porque si no estaría extinta. Por lo tanto, se puede concluir que la esponja no es un contraejemplo porque su sistema nervioso cumple la definición de SPN, incluso si su función computacional es sencilla.

Considerando la cuestión de si el *Trichoplax* es un contraejemplo de la hipótesis de la libertad de movimiento, incluso aunque hay poca información sobre los mecanismos para integrar información, los datos disponibles permiten afirmar que no es un contraejemplo. En base a las investigaciones realizadas, el *Trichoplax* parece que cumple el criterio de un SPN, aunque sus células se enviarían señales químicas y no eléctricas; pero desde un punto de vista computacional, la naturaleza de las señales no es relevante para determinar si es o no un SPN.

A continuación, se van a abordar las tres cuestiones finales que han emergido al comienzo de la sección sobre la hipótesis de la libertad de movimiento. Primero, ¿un sistema nervioso es condición suficiente y necesaria para predecir? En la revisión y discusión que se han realizado, se ha visto que hay dos tipos posibles de sistema nervioso que los organismos multicelulares han adquirido para la evolución: un STN y un SPN. Si un sistema nervioso es un sistema de transmisión, obviamente la respuesta es no. No es necesario ni suficiente porque no procesa información. En el caso de que el sistema nervioso sea un SPN, el SPN sí es condición suficiente para determinar que el sistema nervioso predice. En el proceso de clarificación de la hipótesis de la libertad de movimiento se sustituyó el término “cerebro” por “SPN” y “predicción” con “para procesar información del entorno y generar movimientos para sobrevivir solo en organismos que usan energía interna para producir esos movimientos”. Puesto que el SPN es un sistema de procesamiento y los organismos multicelulares no pueden malgastar energía y ser seleccionados por la naturaleza, el que una especie de organismo posea un SNP es una condición suficiente para determinar que el SNP lleva a cabo predicciones. Sin embargo, el tener un SNP no es una condición necesaria para llevar a cabo predicciones puesto que un SPN no es el único elemento en la clase de los sistemas de procesamiento. Los organismos pueden usar sistemas de procesamiento de diferentes tipos para realizar predicción.

Debería notarse que la hipótesis de la libertad de movimiento no está afirmando que el sistema nervioso es el único mecanismo que puede realizar predicción. Esto es

importante porque uno podría interpretar que los descubrimientos de organismos que pueden predecir sin poseer un SPN son una evidencia en contra de que la principal función del cerebro es la predicción. Esa interpretación sería errónea porque lo que indican los descubrimientos que se han realizado es que la predicción es un proceso más común en la naturaleza de lo que se había pensado hasta ahora. Esto no contradice la hipótesis de la libertad de movimiento porque lo que establece la hipótesis de la libertad de movimiento es la relación entre el SPN y el organismo multicelular con capacidad de moverse.

La siguiente cuestión aborda el movimiento: ¿es la capacidad de moverse una condición necesaria y suficiente para un sistema de procesamiento neuronal? Un organismo multicelular se mueve al menos en dos dimensiones cuando el espacio es heterogéneo, lo que implica que para predecir tiene que analizar señales multidimensionales. Un sistema nervioso que es un SPN es una condición suficiente para generar movimiento y predecir, pero no es una condición necesaria porque la clase de sistemas de procesamiento contiene otros elementos aparte del SPN, como ya se ha mencionado. Sin embargo, hasta el momento, en toda la bibliografía revisada no se ha encontrado un caso de un organismo multicelular que se mueva sin usar un SPN para predecir. Por lo tanto, la hipótesis de la libertad de movimiento parece ser verdad, aunque no se puede explicar por qué. Esto nos devuelve a la principal cuestión de este capítulo: ¿puede la hipótesis de la libertad de movimiento explicar por qué el sistema de procesamiento neuronal es seleccionado frente a otros tipos de sistemas de procesamiento por la naturaleza? La respuesta es que no puede porque el SPN no es una condición necesaria para predecir y moverse en un entorno en el que es necesario procesar señales multidimensionales. Por lo tanto, la hipótesis de la libertad de movimiento no puede explicar por qué el SPN es elegido frente a otro tipo de sistemas de procesamiento.

### 7.2.3. Alternativas a la hipótesis de la libertad de movimiento

Uno podría considerar que, aunque no se ha podido probar que la hipótesis de la libertad de movimiento es falsa, esta podría serlo, por lo que no es adecuada para responder la pregunta principal de este capítulo. Una alternativa que se ha propuesto a la hipótesis de la libertad de movimiento es que la predación es la causa del surgimiento de los sistemas nerviosos [397, 398, 399]. La primera cuestión a notar es que preda está considerada dentro de la hipótesis de la libertad de movimiento porque la localización de la presa implica que la dirección es fundamental para el predador. La siguiente afirmación se realizó para justificar la hipótesis de la predación:

“Thus, predation is possible in ecosystems lacking animals with neurons. Conversely, ecological considerations indicate that neurons cannot exist in ecosystems lacking predation.” [398]p.257

Dado que la hipótesis de la predación asume que la predación es la causa del surgimiento del sistema nervioso en los animales, la primera de las frases implicaría directamente la falsedad de la hipótesis de la predación porque en un análisis lógico de una expresión condicional si el antecedente es falso y la consecuencia es verdad, el condicional es falso. Respecto a la segunda afirmación de que las neuronas no pueden existir en ecosistemas careciendo de predación no es verdad. Un ejemplo, son los tunicados que ya hemos mencionado. Los tunicados tienen un cerebro que solo busca un sustrato adecuado. Además, la *mimosa pudica* tiene potenciales de acción [594, 529] y reacciona a ciertos factores estresantes como el viento y el fuego. Sus respuestas a los factores estresantes la protegen, lo que aumenta su adaptación biológica a su nicho. Claramente, los tunicados y la *mimosa pudica* no son predadores y pueden vivir en entornos sin predadores. Sus sistemas nerviosos han sido seleccionados por la naturaleza porque les aporta una mayor adecuación biológica. El problema de la hipótesis de la predación es que considera solo comportamientos reactivos a cambios relativos a la posición de la presa y omite comportamientos reactivos y proactivos no relacionados con la predación, como detectar y evitar rocas, arrecifes, agujeros, lava, fuego o altas temperaturas y descubrir adecuados lugares para vivir. Sin duda, la predación es un importante elemento y debería ser investigado porque arroja luz sobre la evolución del sistema nervioso; sin embargo, no se puede considerar la única causa. La respuesta al estrés es otro elemento que parece haber jugado un papel importante en la evolución del sistema nervioso [323]. Ser expuesto a estrés o a predación causa que la habilidad de moverse sea seleccionada, ya que esa habilidad permite al organismo evitar la fuente del estrés o el ser predado. Además, a su vez, moverse incrementa o decrementa los niveles de estrés y la habilidad para predear o ser víctima de la predación. Por lo tanto, las causas mencionadas están conectadas, creando un círculo, y cada una de ellas puede iniciar el círculo. Aun así, las causas pueden solo explicar por qué se seleccionan los organismos con un mecanismo para controlar sus movimientos, pero no pueden explicar porque se selecciona un tipo de mecanismo concreto cómo es el SPN frente al resto.

### 7.3. Características de los sistemas nerviosos

Dado que se busca encontrar por qué el SPN es elegido frente a otros tipos de sistemas de procesamiento, las propiedades de neuronas y sistemas nerviosos son importantes para la discusión del sistema nervioso y las revisamos ahora. Esta revisión permitirá llevar una discusión en la sección 7.4 considerando las ventajas y desventajas de los sistemas de procesamiento neuronal frente a otros tipos de sistemas de procesamiento.

### 7.3.1. Características biológicas del sistema nervioso

La función del sistema nervioso de un organismo es procesar y transmitir información, pero algunas veces se olvida que además sirve para generar señales. Por ejemplo, las oscilaciones de las neuronas de oliva inferior parecen ser significantes en la dinámica y sincronización de la organización de las secuencias motoras en la coordinación motora [164]. Una diferencia importante entre neuronas, la unidad biológica del sistema nervioso <sup>1</sup>, y otras células de un organismo multicelular es que las neuronas no pueden dividirse en dos o más células hijas por sus características estructurales, tales como la conformación del árbol dendrítico. Las neuronas, que son creadas desde progenitoras esféricas o células madre, viajan complejos caminos en su migración hasta que alcanzan su posición final. Desde allí, hacen crecer axones y dendritas para establecer conexiones y crear circuitos neuronales. Además, hay que notar que la geometría está conectada directamente a su función, que es la transmisión de información o señales. Por lo tanto, aunque se puedan generar nuevas neuronas, las neuronas no pueden dividirse porque esta acción no es compatible con su función.

A simple vista, la incapacidad de las neuronas para dividirse podría considerarse un serio obstáculo. Sin embargo, la robustez computacional al daño es una característica consolidada del sistema nervioso. Las enfermedades degenerativas del cerebro (como encefalopatía espongiiforme y Alzheimer) han mostrado que el cerebro puede resistir un daño considerable antes de que la función sea seriamente alterada o perdida. La amiloidosis que ocurre en el Alzheimer provoca la inhibición de la transmisión sináptica completamente [405]. A pesar de ello, los primeros síntomas cognitivos de la enfermedad comienzan a manifestarse entre 10 y 15 años después de su comienzo.

Otra característica del sistema nervioso es la plasticidad, lo que puede incluir dos aspectos diferentes. Uno concierne a aprender mientras el otro a conservar la función. En la comunidad científica hay discusión acerca de lo que puede y no puede ser aprendido por el sistema nervioso. Hay muchas opiniones sobre los límites de la plasticidad, y hay pruebas empíricas de que los límites existen [56]. Sin embargo, esta cuestión está más allá del alcance de esta tesis, que se ocupa del segundo, y mucho menos ambicioso, aspecto de la plasticidad. Cuando las neuronas cercanas se hacen cargo de la función de neuronas dañadas de manera que el procesamiento de la información no se ve afectado, el sistema nervioso puede mantener su función. Un ejemplo de este tipo de plasticidad ocurre después de un derrame o un episodio epiléptico en humanos. Además, la plasticidad en el adulto puede ser inducida en respuesta a una lesión o amputación de nervios periféricos [601]. Más adelante se discutirá la diferencia entre plasticidad y robustez.

---

<sup>1</sup>Los astrocitos también juegan un importante papel en las funciones de las neuronas

### 7.3.2. Características computacionales del cerebro neuronal

Las redes neuronales biológicas inspiraron la creación de las redes neuronales artificiales (RRNNAA) [359]. RRNNAA han sido bastante exitosas ejecutando tareas de reconocimiento de patrones. Las RRNNAA pueden aprender patrones (o clases) que rápidamente soportan ruido o variación en los datos de entrada. Sin embargo, uno de los problemas con las RRNNAA en inteligencia artificial es que es difícil recuperar el conocimiento aprendido o almacenado porque está esparcido a través de la RNA. De hecho, recuperar el conocimiento que contiene una RNA continúa siendo un problema importante en IA. Esta cuestión está también relacionada con la robustez al daño porque la dificultad para recuperar información de ellos y su robustez computacional son consecuencia de que las RRNNAA son sistemas distribuidos. Sin embargo, la robustez computacional suele ser siempre pasada por alto porque las RRNNAA son máquinas virtuales. Una máquina virtual es una simulación de un mecanismo computacional llevada a cabo por otro tipo de mecanismos diferente. Esto es posible porque la máquina virtual que ejecuta la simulación tiene igual, o mayor, poder computacional que la máquina simulada. Las RRNNAA son usualmente ejecutadas en máquinas con arquitecturas de registros (e.g. PCs, laptops), pero una máquina de registros no es robusta al daño. Así, si hay un fallo en una de las unidades de la máquina de registros, la máquina fallará también. Esto no ocurriría si la RNA fuera la máquina física. Desde el punto de vista computacional, se han hecho estudios para ver qué ocurre a la función RNA cuando sus nodos o enlaces están dañados [182, 254]. Los resultados de los experimentos han mostrado que el deterioro de la salida es muy pequeño incluso cuando un gran número de nodos o enlaces están dañados. Las redes neuronales no solo son robustas en el procesamiento de información sino también generando señales. Las RRNNAA son usualmente modeladas de una manera sencilla, sin tener en cuenta las propiedades electrofisiológicas intrínsecas. Sin embargo, esas características ya confieren robustez computacional. Además, se ha demostrado que ciertas propiedades electrofisiológicas intrínsecas de las neuronas les permiten auto sincronizarse entre ellas [333]. Esta capacidad de autosincronizarse conlleva robustez al daño también, ya que cuando un organismo pierde parte de sus neuronas o incluso de sus conexiones el sistema puede resincronizarse, generando un nuevo patrón que permite el mantenimiento de una función útil [259].

La robustez computacional al daño de las redes neuronales es una característica diferente de la plasticidad. La robustez computacional es una propiedad inherente del método en el que las redes neuronales procesan información. La plasticidad es un proceso paralelo al procesamiento de información llevado a cabo por la red que altera la estructura y las conexiones del sistema nervioso. Sin embargo, mientras que la robustez computacional permite que la red mantenga su función tras sufrir un daño, la plasticidad permite la generación de una configuración alternativa de procesamiento de información tras un daño. Desde un punto de vista computacional, la robustez es

debida a la propia estructura física para procesar información en las redes neuronales, mientras que la plasticidad requiere de reglas que modifiquen la estructura física con el procesamiento de información.

## 7.4. Sistemas de procesamiento neuronal frente a otros tipos de sistemas de procesamiento

Se ha mostrado que responder la cuestión inicial del capítulo explica a su vez otras cuestiones. La existencia de un SPN en todos los organismos multicelulares con tejido muscular parece revelar un proceso de selección natural. Si esto fuera así, debe haber una ventaja evolutiva que los otros sistemas de procesamiento no poseen y da mayor adecuación biológica a los organismos que tienen un SPN. Se sabe que hay principios energéticos que conducen el proceso de selección natural en general [46, 45] y en los sistemas nerviosos en particular [428]. Sin embargo, esto no puede explicar la selección del SPN por la naturaleza, ya que cualquier sistema de procesamiento puede tener muchas variantes (diferentes tamaños de los elementos de procesamiento, diferentes consumos energéticos, y diferentes formas que poseen valores diferentes de disipación de calor) entre las cuales alguna cumpliría con los principios energéticos. Sin embargo, no se ha encontrado ningún otro tipo de sistema de procesamiento en organismos multicelulares. Así, el SPN debe poseer una característica que dé una ventaja evolutiva a los organismos multicelulares frente a los organismos multicelulares con otros sistemas de procesamiento. Por lo tanto, los principios energéticos pueden explicar la morfología final de un tipo de sistema concreto de procesamiento, pero no pueden explicar la selección del SPN frente a otros sistemas de procesamiento con el mismo poder computacional. Ni axones, ni potenciales de acción, ni dendritas o sinapsis son elementos que dirigen al surgimiento del SPN como la única opción para el procesamiento de información. Ante esta situación la cuestión que debe ser abordada es ¿por qué SPN es seleccionado entre todos los posibles modelos de computación? Para responder a esta cuestión se necesita comparar el SPN con otros tipos de sistemas de procesamiento puesto que en última instancia la selección natural de SPN consiste en por qué un sistema de procesamiento tiene mayor adecuación biológica que otros.

### 7.4.1. Los otros posibles sistemas de procesamiento

Este apartado es un ejercicio de imaginación para proponer sistemas de procesamiento alternativos al SPN. La razón por la que se necesita hacer esto es porque no se conocen sistemas de procesamiento alternativos en la naturaleza. Si han existido en algún momento, no fueron seleccionados frente al SPN. Sin embargo, aunque se trate de un ejercicio de imaginación, los tipos de sistemas de procesamiento deben ser

plausibles con respecto a la biología que se conoce porque se quiere conocer cuál es la ventaja evolutiva que un SPN posee. Por lo tanto, para responder la principal cuestión de este capítulo, los tipos de sistemas de procesamiento con los cuales comparar el SPN deben de haber tenido la oportunidad de existir. La noción de oportunidad es una característica importante a la que usualmente no se le da relevancia que merece en el proceso evolutivo. Esta noción se puede explicar tomando de nuevo la analogía del jugador de dados. Obviamente, cada vez que el jugador lanza el dado, el jugador va a obtener un número. Pero imagínese que el jugador elige un dado entre un grupo de dados, y el grupo es heterogéneo, es decir, hay dados con diferentes números en sus caras. Por lo tanto, dependiendo del dado, hay números que no se podrán obtener en su lanzamiento. Por supuesto, el jugador obtiene un número porque hay una oportunidad de conseguirlo. Sin embargo, no es fácil saber si un número no es obtenido simplemente porque no apareció en el lanzamiento o porque no había oportunidad de que apareciera. Para conocer si un número no es obtenido porque carece de oportunidad se debe conocer qué dado está lanzando el jugador. De la misma manera que un jugador no puede obtener un número si el dado no lo tiene, el universo no puede transmitir eléctricamente información si no existe la fuerza eléctrica. Por lo tanto, el ejercicio de imaginación debe estar de acuerdo a la biología que se conoce que es la que se sabe que es posible en el universo. Además, los sistemas de procesamiento no deben de ser excluidos por los requisitos establecidos por las tareas establecidas por la hipótesis de la libertad de movimiento. Estos requisitos son dos:

1. La habilidad de moverse libremente en la escala macroscópica fuerza a que el organismo debe llevar a cabo complejos procesos de predicción. Por lo tanto, el poder computacional de los sistemas de procesamiento debe alcanzar el límite de Church-Turing<sup>2</sup>.
2. La habilidad de moverse libremente implica la necesidad de enviar señales suficientemente rápido como para coordinar tejido contráctil en el cuerpo del organismo. Sin coordinación, el organismo no puede generar un comportamiento efectivo para alcanzar una meta. Así, si el organismo tiene tejido contráctil, debe de coordinarlo y controlarlo durante ese periodo de tiempo.

Tres posibles sistemas de procesamiento cumplen las condiciones que se han establecido y pueden tener la posibilidad de existir de acuerdo con la biología que conocemos actualmente. Estos sistemas son los siguientes: la superneurona, el sistema P sináptico y la red de Petri.

---

<sup>2</sup>Obviamente, no se considera que el sistema tenga que tener una cantidad infinita de memoria

#### 7.4.1.1. La superneurona

Una superneurona sería una célula cuyo árbol dendrítico podría llevar a cabo toda la computación realizada por un grupo de neuronas. Este modelo está basado en hechos biológicos: hay dendritas que poseen axones, las dendritas tienen potenciales y las células de Betz. El primer hecho mencionado hace referencia a que recientemente se ha descubierto que existen células piramidales del hipocampo que tienen dendritas basales de las que emanan axones [569]. El segundo hecho es que las dendritas poseen potenciales de acción [313], determina que les permite llevar a cabo su propio procesamiento de información. El tercero que menciona a las células de Betz, es debido a que son células piramidales gigantes localizadas en el córtex motor primario que envían sus axones a la medula espinal. Así su axón puede alcanzar hasta un metro de longitud. Adicionalmente, se puede citar que las jirafas tienen axones que poseen varios metros de longitud recorriendo la longitud completa de sus cuellos. Así, no es extravagante considerar una neurona que tenga dendritas con axones que hacen sinapsis con otras dendritas de la misma neurona, repitiéndose estas concesiones a través del árbol dendrítico. Así, la superneurona podría llevar a cabo todo el procesamiento realizado por un grupo de neuronas. Obviamente, no es factible que una superneurona lleve a cabo el procesamiento de todo el cerebro humano, pero podría realizar el procesamiento de un ganglio de un invertebrado. Además, se podría imaginar la existencia de cerebros hechos de superneuronas.

#### 7.4.1.2. Sistema P sináptico

La computación molecular es un método usado por microorganismos para procesar información [74]. El poder de la computación molecular ha sido estudiado durante las últimas décadas [318, 226, 331, 93, 540, 474, 434]. La computación molecular se divide fundamentalmente en dos campos: computación mediante ADN y computación con membrana. Un sistema-P es un modelo computacional que realiza cálculos usando procesos inspirados por la biología [446]. Se trata de una representación del modo en que sustancias químicas interactúan y cruzan la membrana celular. De acuerdo a las investigaciones realizadas, se propone que un sistema de procesamiento que podría competir con el SPN es el sistema P sináptico. Un sistema P sináptico sería una célula con la capacidad de procesar información que tiene un grupo de neuronas llevada a cabo a través de la computación molecular. La plausibilidad de este modelo está basada en la existencia de receptores acoplados a proteínas G. Las proteínas G permiten a las neuronas recibir señales de entrada y activar enzimas que disparan reacciones bioquímicas, que modifican la función neuronal. Esas reacciones químicas son llamadas cascadas de segundo mensajero. Basándose en lo anterior, se puede considerar posible que las cascadas de segundo mensajero podrían implementar un mecanismo para procesar información en el que diferentes tipos de proteínas G permitirían diferentes

valores de entrada. Por lo tanto, parece plausible que un sistema P sináptico tuviera la oportunidad de existir. Los sistemas P sinápticos no deben ser confundidos con una extensión de los sistemas P denominada sistema P neural de pulsos [242]. Los objetos en un sistema P neural de pulsos evoluciona por medio de las reglas de pulsos mientras que en un sistema P evoluciona por medio de interacciones químicas. Como en el caso de las superneuronas, un sistema P sináptico podría llevar a cabo el procesamiento que realiza un ganglio y podrían existir cerebros constituidos por sistemas P sinápticos.

#### 7.4.1.3. Red de Petri

Una red de Petri es un modelo computacional muy conocido inventado por Carl Adam Petri [458]. Las redes de Petri pueden computar elecciones, interacción, y ejecutar concurrencia, y se ha demostrado que una red de Petri a la que se le permiten arcos inhibitorios puede simular una máquina de Turing [11]. Se pueden usar para producir datos cuantitativos calculando ecuaciones diferenciales parciales [55]. Una red de Petri contiene lugares, transiciones, arcos y fichas. Los arcos conectan lugares, pero siempre con una transición entre los lugares conectados. Las fichas se mueven a través de la red. Los estados de una red de Petri son definidos por los lugares donde las fichas están situadas. Una red de Petri podría ser creada por la naturaleza de la siguiente manera: lugares y transiciones serían células, denominadas células lugar y células transición. Las células lugar serían células que reciben sinapsis desde las células receptoras o las células transición. Cada sinapsis activada en una célula lugar implicaría la activación de proteínas G. Las proteínas G se activarían para que se produjera un potencial de acción y también controlaría la probabilidad de liberación y transmisión. La liberación ocurriría cuando un umbral de alguna enzima es alcanzado. La concentración de la enzima se incrementaría con la activación de las células. Así, se reflejaría la acumulación de fichas en los lugares. Cada célula lugar tendrá un estado umbral en el cual un potencial de acción se disparará. Las células lugares podrían formar sinapsis con las células transición. Las células transición registrarían si la condición es cumplida por las células lugar que convergen en ellas mediante un mecanismo que existe en las neuronas para reconocer la coincidencia de señales de entrada en el tiempo. El mecanismo fue descubierto en las neuronas piramidales de la capa 5. Estas piramidales tienen diferentes estados de potencial en reacción a acoplamiento de señales [288]. Por lo tanto, las células podrían tener diferentes estados potenciales. En cada estado, canales voltaje dependientes diferentes estarían trabajando. La sinapsis entre células lugar y células transición modificarán el estado potencial de la célula transición. Cuando una célula de transición en el estado potencial más positivo recibe el disparo de una sinapsis, la célula transición disparará un potencial de acción para activar sus sinapsis con las células lugar, y el estado potencial cambiará al estado potencial más negativo. Si el estado potencial para generar potenciales de acción en una célula transición sería el más positivo o negativo dependería de cuál es el modo más eficiente energéticamente.

Por todo lo mencionado parece plausible que existe la oportunidad de que una red de Petri pudiera existir en la naturaleza como sistema de procesamiento para organismos multicelulares.

#### 7.4.2. Selección natural del sistema de procesamiento neuronal

En el apartado 7.2.2, se ha visto que la habilidad de moverse implica la necesidad de un sistema potente de computación para un organismo multicelular pero no explica por qué el tipo de sistema procesamiento seleccionado por la naturaleza es el SPN. Así, si la selección natural escoge el modelo de red neuronal, debe de haber al menos una ventaja evolutiva que diferencia el modelo de red neuronal del resto de modelos computacionales. Por lo tanto, si se quiere responder a la principal pregunta de este capítulo será necesario responder a la pregunta de ¿Cuál es la ventaja evolutiva por la que SPN es seleccionado frente a otros tipos de sistemas de procesamiento? Considerando distintos puntos de vista, al menos siete hipótesis pueden ser formuladas sobre la causa de la selección del SPN. Las hipótesis que se contemplan son las siguientes:

- Hipótesis F (Fastest): SPN es seleccionado frente a otros tipos de sistemas de procesamiento porque SPN es el sistema de procesamiento más veloz debido a que procesa en paralelo.
- Hipótesis M (Maximum): SPN es seleccionado frente a otros tipos de sistemas de procesamiento porque SPN es el que provee el máximo poder computacional.
- Hipótesis N (Noise): SPN es seleccionado frente a otros tipos de sistemas de procesamiento porque SPN puede soportar ruido en las señales de entrada mejor que el resto.
- Hipótesis O (Optimum): SPN es seleccionado frente a otros tipos de sistemas de procesamiento porque SPN es más eficiente energéticamente que el resto.
- Hipótesis P (Plasticity): SPN es seleccionado frente a otros tipos de sistemas de procesamiento porque SPN tiene más plasticidad para recuperar su función que el resto.
- Hipótesis R (Robustness): SPN es seleccionado frente a otros tipos de sistemas de procesamiento porque SPN tiene mayor robustez computacional al daño que el resto.
- Hipótesis V (Variety): SPN es seleccionado frente a otros tipos de sistemas de procesamiento porque SPN permite mayor variabilidad de métodos computacionales que el resto.

El siguiente paso es analizar cada una de estas hipótesis cómo una explicación a por qué el SPN es seleccionado.

#### 7.4.2.1. Hipótesis F

La hipótesis F propone que la selección del SPN habría sido debida a que el SPN procesa información más rápido que cualquier otro tipo de sistema de procesamiento dado que el SPN es un mecanismo que procesa en paralelo. Sin embargo, el pensamiento humano es secuencial [637], y cualquier tipo de procesamiento simbólico hecho por animales es secuencial. Por lo tanto, el procesamiento paralelo del SPN ralentiza el procesamiento secuencial, ya que sería más rápido con un sistema puro de procesamiento secuencial. Además, la superneurona, el sistema P sináptico y la red de Petri pueden realizar procesamiento en paralelo. Así, la hipótesis F no es la razón por la que la naturaleza ha seleccionado el SPN.

#### 7.4.2.2. Hipótesis M

La hipótesis M argumenta que el SPN debería ser seleccionado debido a que posee el máximo poder computacional entre todos los sistemas de procesamiento. Una red neuronal consigue más poder cuando su número de capas es incrementada [492]. Sin embargo, el incremento tiene un límite superior, que se conoce como el límite de Church-Turing. Las redes neuronales alcanzan ese límite<sup>3</sup>, pero la red de Petri, la superneurona y el sistema P sináptico lo alcanzan también. Por lo tanto, el poder computacional no puede ser la causa de la selección del NPS.

#### 7.4.2.3. Hipótesis N

La hipótesis N referencia dos tipos de ruidos diferentes: el ruido exterior y el ruido interior. El ruido exterior no puede ser la causa porque una superneurona podría lidiar con él de la misma manera que el SPN, y la red de Petri podría usar algoritmos secuenciales que se conocen para tratar el ruido. Respecto al ruido interno que se provocaría por el fallo de la transmisión sináptica del cuanto, pero esto no es un problema intrínseco del sistema nervioso porque hay observaciones fuera del neocórtex de transmisión sináptica de cuantos sin fallo [445, 51]. Así, la existencia de fallos en la transmisión sináptica de cuantos existe debido a que no provocan un perjuicio al organismo, ya que la naturaleza podría evitar el fallo si fuera perjudicial para el organismo. De hecho, se ha demostrado que los fallos en la transmisión sináptica de cuantos permite una mayor eficiencia energética en el procesamiento de información [300]. Incluso, los datos obtenidos en experimentos han mostrado que los terminales sinápticos pueden

---

<sup>3</sup>Los modelos de red neuronal que superan el límite Church-Turing no son considerados porque no hay evidencia experimental de que esto ocurra en la naturaleza [530].

cambiar individualmente la probabilidad de liberación de neurotransmisores dinámicamente [73]. Por lo tanto, SPN no puede haber sido seleccionado para soportar su ruido interno durante la computación porque el ruido es una característica para mejorar la eficiencia energética del SPN en el procesamiento de información.

#### 7.4.2.4. Hipótesis O

La hipótesis O propone que la selección del SPN sería causada porque NPS es el tipo de sistema de procesamiento que gasta menos energía. Se sabe que la optimización de energía tiene un importante papel en la evolución del sistema nervioso. A pesar de ese hecho, todos los estudios muestran que los sistemas nerviosos son muy caros en coste energético. Incluso así, uno podría preguntarse si es el menos caro entre todos. La superneurona sería en el peor de los casos igual de costosa que una red neuronal, y en el mejor de los casos sería menos costosa porque la maquinaria estaría más concentrada, ahorrándose en estructuras. En el caso de la red de Petri, esta gastaría menos energía que el SPN porque tendría menos unidades ya que no realiza un procesamiento paralelo redundante. El sistema P sináptico debería ser menos costoso energéticamente que el SPN. Esto es así porque en el caso del sistema P sináptico la computación se produce a una escala molecular mientras que la computación en el SPN crece al menos linealmente con el número de unidades.

#### 7.4.2.5. Hipótesis P

La hipótesis P parece plausible a primera vista, pero el hecho de que la plasticidad puede restituir la función de un sistema tras un daño si el sistema de procesamiento es robusto ante el daño excluye la hipótesis P de ser la causa de que el SPN haya sido seleccionado. Si el sistema de procesamiento no es computacionalmente robusto y cualquier daño provoca un fallo total del sistema, la plasticidad es inútil. Si una superneurona tuviera un proceso de necrosis o apoptosis, aunque la superneurona tenga plasticidad, no es posible que la superneurona continúe procesando información porque tras la necrosis o apoptosis está destruida. Por lo tanto, esta conclusión apunta a la hipótesis R.

#### 7.4.2.6. Hipótesis R

La hipótesis R explica que la selección de la SPN es debida a una característica especial que las redes neuronales poseen, la robustez computacional al daño. Usualmente, la característica que es resaltada de las redes neuronales en el campo de la informática es su flexibilidad ante los datos de entrada, mientras que la robustez computacional al daño es mayormente ignorada. Esto se debe a que las redes neuronales son simuladas sobre otros sistemas computacionales que no poseen robustez computacional. Sin embargo, esta característica si es poseída por los SNP.

#### 7.4.2.7. Hipótesis V

En el capítulo anterior se ha mostrado que las redes neuronales permiten diferentes métodos de procesamiento de información. Así, se podría considerar que la variabilidad que permite el modelo de red neuronal junto con la variabilidad genética podría permitir que aparecieran individuos cuya arquitectura neuronal se adaptase mejor a los requisitos computacionales del entorno. Sin embargo, cualquier modelo computacional que alcance el poder de la máquina de Turing puede implementar cuantos algoritmos haya para codificar información.

#### 7.4.2.8. La robusted computacional es una ventaja evolutiva

El análisis llevado a cabo de las hipótesis ha mostrado que la robustez computacional es la única característica de las 7 que diferencia el SPN de los otros sistemas de procesamiento. Ahora, la cuestión a analizar es si esta característica es una ventaja evolutiva. Lo primero a mencionar es que la robustez computacional no puede ser ignorada en sistemas biológicos porque cuanto mayor es el periodo de tiempo que esta vivo un organismo, mayor es su exposición al daño, y el daño en el sistema nervioso tiene un resultado fatal para el organismo [258]. Un organismo multicelular es caracterizado por la especialización de sus células. Mientras que ese hecho implica un mejor rendimiento en su función, también implica interdependencia para sobrevivir. El fallo de un órgano que se encarga de una función provoca la muerte del organismo. Desde el punto de vista biológico, la pérdida de células por apoptosis y necrosis surge como un problema que tiene que ser encarado por los organismos multicelulares. El envejecimiento debería implicar al menos una pérdida lineal de células en el tejido. Por ello, los organismos multicelulares tienen procesos de regeneración de crecimiento compensatorio para solventar el problema de las pérdidas de células. Sin embargo, los organismos multicelulares no pueden usar esos procesos en la mayor parte de su sistema de procesamiento de información. La especialización de una célula para procesar información para un organismo multicelular produce la pérdida de la capacidad de división. Hay tres razones por las que sucede la pérdida de la capacidad:

- Un organismo debe tomar decisiones constantemente. En un organismo multicelular, el sistema de procesamiento debe coordinar la función de diferentes órganos y sistemas constantemente, y los peligros pueden aparecer en el entorno aleatoriamente. Durante la división, sería imposible para la célula procesar información debido a los movimientos y la rotura de la membrana celular. No es factible que la pérdida de células se compense en el sistema de procesamiento de información por división porque esto implicaría que las células no podrían procesar información durante la división, por lo tanto, el organismo se quedaría sin la habilidad para responder a cualquier cambio del entorno.

- La compleja forma de la célula para transmitir y procesar información en un organismo multicelular hace la división imposible.
- El procesamiento de información no es una función aditiva. Hay pocos pasos donde es posible añadir nuevas células sin alterar la función computacional que se lleva a cabo. El resto de los órganos de un organismo multicelular tiene una función aditiva: Cuanto mayor es el número de células, mayor la capacidad de producir o procesar del órgano. Sin embargo, un sistema de procesamiento de información no se trata de una estructura con una función aditiva; cada unidad depende específicamente de las otras para que se pueda llevar a cabo la función. Los métodos para establecer conexiones y rutas de procesamiento de información contienen acciones muy complejas para guiar las células por medio de señales químicas. Una vez las conexiones son establecidas, la migración de células para alcanzar lugares específicos necesita ser detenida, ya que en la mayoría de los casos otra etapa comienza en la que las señales químicas de la etapa anterior deben ser detenidas. Por lo tanto, el proceso de añadir nuevos elementos al sistema de procesamiento como ocurre en otros sistemas es inviable porque añadir nuevas unidades modificaría la conectividad, lo que supone un procesamiento erróneo de la información. Solo en los casos donde añadir unidades al sistema no modifique la función serían compatibles con la generación de nuevas neuronas. Por ellos, pocos procesos de neurogénesis existen.

Por lo tanto, hay dos objetivos que entran en conflicto en los organismos multicelulares. Por un lado, un organismo multicelular debe especializarse y crear un sistema para coordinar elementos distantes para generar comportamiento. Por otro lado, el organismo debe mantener todos los sistemas durante la vida del organismo por su interdependencia hasta que logre tener descendencia. Sin embargo, el objetivo no puede ser logrado mediante regeneración, o crecimiento compensatorio. Este dilema implica que hay una presión para seleccionar un tipo de sistema de procesamiento que pueda ser capaz de mantener su función hasta que el organismo tenga descendencia. La superneurona y el sistema P sináptico cesarán su función de procesar información si sufren apoptosis o necrosis. Para un organismo multicelular, esto conllevaría la muerte del organismo porque la superneurona y el sistema P sináptico son sistemas de procesamiento de todo o nada. En el caso de la red de Petri, la apoptosis o necrosis de una de las células no causa el cese del procesamiento de información; pero causaría el bloqueo de la red o modificaría completamente la función computacional que estaba llevando a cabo. Así, el SPN es el único sistema de procesamiento que tiene robustez computacional frente al daño, y la robustez computacional al daño es una ventaja evolutiva.

### 7.4.3. Principios del procesamiento neuronal

La hipótesis R explica sólidamente por qué el SPN es seleccionado frente a otros posibles tipos de sistemas de procesamiento en organismos multicelulares. Usando el descubrimiento de que la robustez computacional es una ventaja evolutiva en los organismos multicelulares que necesitan procesar información, se formula el siguiente principio:

*Primer principio de procesamiento neuronal: En los organismos multicelulares que requieren procesar información de su entorno para sobrevivir, el sistema de procesamiento neuronal es seleccionado frente a otros tipos de sistemas de procesamiento que carezcan de robustez computacional al daño.*

Analizando detalladamente el principio enunciado surge inmediatamente a la siguiente cuestión: ¿Cómo de computacionalmente robusto debe de ser un organismo? La cantidad total de neuronas de un sistema de procesamiento neuronal se puede dividir en dos cantidades. La primera cantidad es el número de neuronas que permiten llevar a cabo una función, y la segunda es la cantidad de neuronas que permite un cierto grado de robustez computacional. El número de neuronas requerido para llevar una función es fijo, pero el otro parámetro es libre. Así, la robustez computacional de una red neuronal depende cuantas neuronas contiene el sistema de procesamiento neuronal; esto es, cuanto mayor es el número de neuronas, mayor es el número que puede perder sin que eso afecte a la función. La cuestión de cómo de robusto computacionalmente debe de ser el SPN de un organismo puede ser también formulada como: ¿Cuántas neuronas debe poseer un SPN? Se sabe que el número no es aleatorio porque el tamaño del cerebro se conserva en las especies, así que, si la hipótesis R es correcta, la naturaleza selecciona el número de neuronas para soportar una cantidad concreta de daño, además de para realizar una función computacional. La respuesta a la pregunta planteada está relacionada con el núcleo de la teoría de la evolución formulada por Charles Darwin. Darwin notó que la descendencia, en las etapas tempranas de existencia, es siempre más numerosa que el número de sus progenitores [238]. Teniendo un número de descendientes mayor que sus progenitores cumple un doble objetivo: permitir que alguno de los descendientes logre una mayor adecuación a través de la variación y amortiguar el impacto de predadores o eventos reduciendo la población. Contestando la cuestión sobre cuantas neuronas de NPS de un organismo debe poseer, otro principio de procesamiento neuronal puede ser formulado tomando en cuenta el hecho notado por Darwin.

*Segundo principio de procesamiento neuronal: En los organismos multicelulares cuyos nichos requieren un sistema de procesamiento neuronal para estar adaptados biológicamente, su sistema de procesamiento debe durar tanto tiempo como necesite el organismo*

*para tener la suficiente descendencia que permita variación.*

Leyendo este segundo principio, uno puede preguntarse a que cantidad se refiere con “suficiente descendencia para permitir variación”. Como se ha mencionado, el número de descendencia debe ser mayor que el número de progenitores, pero el número concreto depende en cada caso de la probabilidad de los descendientes directos para llegar a ser adulto y reproducirse exitosamente.

Sin embargo, debería notarse que la pregunta “¿Qué robustez computacional debe poseer el SPN de un organismo?” y “¿Cuántas neuronas debe poseer un SPN?” no son exactamente la misma cuestión. La diferencia reside en que, aunque el segundo principio responde completamente a la primera cuestión, solo responde parcialmente a la pregunta de cuántas neuronas debe poseer. El segundo principio solo determina la mínima cantidad requerida ya que la selección darwiniana es un proceso de dos pasos [351]. El primer paso es la producción de organismos entre los que hay una amplia cantidad de variación, a lo que le sigue el segundo paso: la selección o eliminación. Inicialmente, todas las opciones que superan la cantidad establecida por el segundo principio podrían ser seleccionadas por el segundo paso. Sin embargo, es posible que haya también un límite máximo al número de neuronas que puede poseer un SPN. Para responder a la cuestión, se ha de descubrir si hay principios que limiten el número máximo de neuronas que un SPN puede poseer. La respuesta es: sí, lo hay. Para alcanzar esta respuesta empezamos con una ley fundamental de la física, la ley de la conservación de la energía. Esta ley implica que la energía no puede ser creada, por lo que el organismo debe tomarla de su entorno. Además de esto, la escasez de energía en la naturaleza conduce a la competición por ella. Por lo tanto, la media diaria de energía consumida por el SPN de un organismo debe ser menor que la media de energía diaria que puede ser obtenida por el organismo en su nicho. La energía tiene que ser estrictamente menor porque el resto de sistemas del organismo necesitan consumir energía también. Tomando todo esto en cuenta, el siguiente principio surge:

*Tercer principio de procesamiento neuronal: En los organismos multicelulares que viven en entornos cuyos nichos requieren un sistema de procesamiento neuronal para estar adaptados biológicamente, el número de neuronas en el sistema de procesamiento no puede exceder el número de neuronas que gastan durante un periodo de tiempo la energía igual a la que puede ser obtenida en ese mismo periodo de tiempo del entorno.*

Se sabe que el tamaño determina la morfología, fisiología, y la biología de las especies [509]. Así, el máximo establecido por el tercer principio no es siempre lograble, ya que un nicho puede entrañar un tamaño para el organismo que implique límites inferiores al máximo del tercer principio. De hecho, las tendencias estructurales asociadas

con la miniaturización han sido descritas en muchos animales [205]. Por lo tanto, otro principio de procesamiento neuronal que limita el número máximo de neuronas puede ser enunciado:

*Cuarto principio de procesamiento neuronal: En los organismos multicelulares que viven en entornos cuyos nichos requieren un sistema de procesamiento neuronal para estar adaptados biológicamente, el número de neuronas en el sistema de procesamiento neuronal no puede exceder el número de neuronas que el volumen del cuerpo del organismo (determinado por su nicho) permite.*

Aunque el cuarto principio parece bastante obvio, este puede explicar el tamaño de las neuronas si se combina con el segundo principio. Así, pueden explicar los descubrimientos hechos sobre las diferentes estrategias desarrolladas por los organismos pequeños para empaquetar las neuronas necesarias en cuerpos de pequeño tamaño. Para cumplir el segundo y cuarto principio al mismo tiempo, la evolución ha encontrado diferentes soluciones. Una solución ha sido conseguir los mayores valores de compresión de cromatina, resultando en una reducción del volumen del núcleo [467]. Otra ha sido generando neuronas anucleadas [468].

Estos cuatro principios implican que los organismos no serán seleccionados por la naturaleza a menos que los cumplan, pero que se cumplan no garantiza su selección.

Debe notarse que, aunque el análisis que ha llevado a los resultados fue hecho usando modelos computacionales específicos, el resultado es general porque cubre todos los seres multicelulares que necesitan procesamiento de información. La generalidad del resultado viene de dos hechos. Primero, un teorema o principio puede ser probado falso con solo un contraejemplo, no se necesita una prueba general. Por lo tanto, eso ha permitido excluir 5 de las 6 hipótesis. Segundo, si la persistencia del modelo de red neuronal ha sido causada por deriva aleatoria, no debería ser posible encontrar una ventaja evolutiva. Sin embargo, se ha encontrado que la robustez computacional al daño es una ventaja evolutiva porque da una ventaja reproductiva frente a tres modelos computacionales biológicamente plausibles.

Tras esta clarificación, ¿se puede responder a la pregunta principal de este capítulo? Se debe decir que hay una respuesta, pero que esta es incompleta. Tras el análisis no se puede explicar por qué el modelo de red neuronal es seleccionado frente a cualquier modelo computacional, ya que en el análisis realizado no se ha considerado ningún modelo computacional que sea biológicamente plausible con robustez computacional al daño. Por lo tanto, se plantean las siguientes preguntas:

- ¿Existen otros modelos computacionales biológicamente plausibles con robustez computacional al daño?
- Si los hay, ¿por qué el modelo de red neuronal ha sido elegido frente a ellos?

Con respecto a la primera cuestión, no puedo imaginar ningún modelo computacional biológicamente plausible con robustez computacional al daño diferente del modelo de red neuronal para organismos multicelulares. Hay una gran probabilidad de que no exista otro modelo computacional biológicamente plausible con robustez computacional al daño. Si este fuera el caso, se tendría una respuesta completa a la principal cuestión, y el primer principio se reformularía de una manera más fuerte:

Primer principio del procesamiento neuronal (fuerte): En cualquier organismo que requiera procesar información de su entorno para sobrevivir, el sistema de procesamiento neuronal es seleccionado frente a cualquier otro tipo de sistema de procesamiento porque posee robustez computacional al daño.

Pero, esta formulación fuerte es solo una hipótesis porque no puedo aportar ninguna prueba de la no existencia de otros modelos computacionales biológicamente plausibles con robustez computacional al daño. En el caso opuesto, se debería al menos poder encontrar otro modelo computacional con robustez computacional al daño con el que estudiar cuál es la ventaja evolutiva que causa la selección del modelo de red neuronal frente a los otros modelos computacionales biológicamente plausible con robustez computacional al daño.

## 7.5. Análisis de los resultados

La informática teórica muestra que hay muchos modelos computacionales diferentes. Además, el proceso de la selección natural tiene la variación como uno de sus elementos necesarios. Si se tienen en cuenta esos dos hechos no es posible que nunca haya emergido otro modelo computacional para el sistema nervioso de organismos multicelulares. Por lo tanto, es lógico preguntarse por qué la red neuronal ha emergido y persistido frente a otros posibles modelos computacionales en la evolución de los organismos multicelulares. Sin embargo, la comunidad científica ha asumido el hecho de que el modelo de red neuronal es el modelo computacional que la naturaleza ha elegido para generar los comportamientos para organismos multicelulares, y mientras la evolución de los sistemas nerviosos es un campo en el que trabajan gran número de investigadores, la cuestión de por qué el modelo de red neuronal ha persistido frente a otros modelos computacionales durante la evolución de organismos multicelulares no ha sido planteada. Este capítulo se ha centrado en abordar esa cuestión porque sin una respuesta a esa pregunta no se puede tener una visión completa de la evolución del sistema nervioso. Además, hay tres preguntas que están íntimamente relacionadas con por qué la naturaleza ha evolucionado usando los SSPN: (1) ¿Ha sido la selección natural o la deriva aleatoria la que ha causado que todos los metazoos tengan un sistema nervioso con el modelo computacional de red neuronal?, (2) ¿existen principios genera-

les en la naturaleza que seleccionan un tipo de organismo en base a las características de sus mecanismos de toma de decisiones? y (3) si es así, ¿cuáles son esos principios? El abordar estas tres preguntas ha mostrado que la cuestión inicial es fundamental para contestar a otras preguntas. Esto da más valor a la pregunta inicial.

Inicialmente, se ha excluido la deriva aleatoria en base a dos hechos. [297, 89]. Así, es difícil que ocurriera un proceso de deriva aleatoria que hubiera provocado la existencia únicamente del modelo de red neuronal. Esto sugiere que habría una ventaja evolutiva que logran los organismos multicelulares al usar el modelo de red neuronal para procesar información. Segundo, la aparición de la multicelular da compleja ha ocurrido varias veces en la naturaleza [426, 268], así como la aparición de neuronas [408, 407]. Sin embargo, solo se puede encontrar el modelo de red neuronal en los organismos multicelulares que necesitan procesar información, lo que indica también que el modelo de red neuronal debe aportar una ventaja evolutiva. Si la interpretación de los indicios fuera equivocada al excluir la deriva aleatoria, y la persistencia del modelo de red neuronal fuera un hecho causado por ella, entonces no se encontraría que el modelo de red neuronal aporta una ventaja evolutiva a los organismos multicelulares. Por otro lado, si estamos en lo correcto, debería de encontrarse una ventaja evolutiva debida al modelo de red neuronal. Por lo tanto, la cuestión de qué ventaja evolutiva posee el modelo de red neuronal para los organismos multicelulares emerge como una cuestión clave para contestar sobre la persistencia de este modelo en la naturaleza.

Antes de discutir esa cuestión, una estructura de clasificación basada en el concepto de procesamiento de información ha sido propuesta para comparar sistemas nerviosos. La clasificación propuesta tiene la ventaja de no estar restringida por las características anatómicas pertenecientes a una especie, y se adapta a diferentes grados de especificación.

La investigación para responder la cuestión principal comenzó con una propuesta hace 30 años sobre la función del cerebro y su origen [314, 315]. La propuesta establece que la predicción es la principal función del cerebro y que todos los organismos multicelulares que se mueven requieren un cerebro (un sistema nervioso) para hacer predicciones. Esta propuesta ha sido denominada la hipótesis de la libertad de movimiento. La formulación inicial de la hipótesis de la libertad de movimiento puede dar lugar a ciertas ambigüedades, así que se ha realizado una reformulación para evitar ambigüedades, seguida por una discusión sobre la verdad de la afirmación y si puede responder a la cuestión principal que aborda el capítulo. Aunque nada descubierto durante los últimos 30 años refuta la hipótesis de la libertad de movimiento, no puede responder la pregunta principal de este capítulo. Otras causas alternativas, como la hipótesis del estrés y de la predación, han sido consideradas también, pero, ninguna de ellas responde las cuestiones. Por lo tanto, la cuestión de por qué de la persistencia del modelo computacional frente a otros modelos computacionales en los organismos multicelulares ha tomado una mayor importancia.

Con el objetivo de intentar resolver la cuestión, se han propuesto siete hipótesis

como posibles respuestas a la cuestión inicial, y cada una fue analizada considerando si explica o no por qué el modelo de red neuronal es el único que se encuentra en los organismos multicelulares para procesar información. La conclusión final de este análisis es que la robustez computacional al daño es una ventaja evolutiva que ha condicionado la evolución del sistema de procesamiento para generar comportamiento en los organismos multicelulares. La robustez computacional al daño permite una mayor adecuación biológica porque los organismos multicelulares son capaces de mantener su función a pesar de los daños. Usando esta conclusión, se ha formulado el primer principio de procesamiento neuronal.

El primer principio de procesamiento neuronal afirma que para los organismos multicelulares el modelo de red neuronal es seleccionado para los sistemas nerviosos frente a otros sistemas de procesamiento porque tiene robustez computacional que le permite mantener su función al sufrir daños mejor que otros modelos computacionales de media. Sin embargo, el primer principio abre otra pregunta: ¿Cuánto daño debe de ser capaz de soportar el SPN de un organismo multicelular? Esta cuestión aparece porque existe un parámetro que está libre en el primer principio, y ello está relacionado con preguntar cuántas neuronas debe poseer un SPN. Para responder a esa pregunta, se ha formulado un segundo principio usando uno de los hechos fundamentales en los que se basó Darwin para elaborar su teoría de la evolución. Ese segundo principio afirma que el SPN de un organismo debe durar el tiempo necesario para asegurar que su descendencia directa sea más numerosa que el número de progenitores. Aunque el segundo principio responde la cuestión del grado de robustez computacional que es necesario, solo responde parcialmente a la pregunta de cuántas neuronas debe poseer un SPN porque solo determina el mínimo. Así, se ha buscado determinar el máximo. Se ha visto que hay un máximo determinado por las restricciones energéticas del entorno, aunque otros tipos de restricciones, como el tamaño máximo del cuerpo permitido por el nicho, pueden obligar a que el máximo sea menor. En base a lo anterior se han enunciado dos principios más: el tercer y cuarto principio de procesamiento neuronal. Las cuatro afirmaciones consideradas en los principios parecen simples porque su verdad parece evidente, pero permiten la explicación de hechos complejos de la evolución de los sistemas nerviosos.

Hasta ahora la robustez computacional ha carecido de atención en el campo evolutivo, pese a que en investigaciones sobre el cerebro humano si se ha considerado esta característica [412, 88, 16]. Sin embargo, los principios de procesamiento neuronal presentados en este capítulo resaltan la importancia de la robustez computacional y la minimización del daño. Tras las conclusiones obtenidas de la investigación, la robustez computacional es una característica esencial para que los organismos puedan ocupar muchos nichos de la naturaleza. Durante las últimas décadas, los requisitos energéticos han sido un tema importante porque se ha mostrado que son un parámetro importante que dirige la evolución en general [45, 46] y la del sistema nervioso en particular [428]. Sin embargo, hay preguntas que los requisitos energéticos no pueden explicar, tales

como por qué la naturaleza elige solo un modelo de computación para el reino de los Eumetazoos.

La discusión de la cuestión inicial de este capítulo resalta que la robustez computacional al daño y la minimización del daño son cuestiones claves en múltiples pasos de la evolución. Hasta ahora, la robustez computacional al daño para los sistemas de procesamiento de información no ha sido discutida desde una perspectiva evolutiva. Los resultados presentados en esta tesis muestran que explorar otros parámetros además de los energéticos permite alcanzar una visión más completa del proceso evolutivo.

## Capítulo 8

# La TGE y el paradigma computacional

Las ciencias cognitivas, en especial la psicología cognitiva y la IA, propone explicar la mente como un procesador de información, y usan el paradigma computacional como marco para abordar la explicación de la mente. La TGE evita cualquier referencia a la formulación mentalista para explicar los comportamientos, pero coincide con la ciencia cognitiva en la formulación de que los comportamientos son generados por procesos computacionales. La TGE propone que la naturaleza de la fasa es computacional, y, por lo tanto, sus valores se describen computacionalmente. Esto implica que, si existieran limitaciones en el paradigma computacional, estas afectarían también a la TGE. Así, cabe preguntarse, ¿es capaz el paradigma computacional de describir todos los valores de fasa? Como se comentó en el capítulo 2, durante la primera parte del siglo XX, dos enfoques muy diferentes estuvieron luchando para explicar los comportamientos de humanos y otros animales en la psicología. Uno fue el conductismo, que estaba enfocado en el estudio del comportamiento y había logrado algunos éxitos. Este enfoque estaba fuertemente establecido en Estados Unidos. El otro enfoque fue una perspectiva del desarrollo. La perspectiva del desarrollo nació en Europa y se ocupó especialmente de las representaciones mentales [591]. En los años 50, los experimentos empezaron a hacer evidente que el conductismo no era capaz de explicar diversos fenómenos y que era necesario el uso de teorías para describir estados internos. Al mismo tiempo, algunos descubrimientos en psicología mostraron que algunos procesos mentales eran explicables como procesos computacionales [430]. El resultado fue el comienzo de las ciencias cognitivas. Las ciencias cognitivas postulan que la mente es una estructura de datos y la cognición es un proceso computacional [382]. Los trabajos de Alan Turing y W.S. McCulloch, fueron la base sobre las que los investigadores comenzaron a desarrollar las ciencias cognitivas para explicar cómo funciona la mente. Recuérdese que se mencionó que en los primeros momentos de la IA, Newell y Simon plantearon que los programas podían ser considerados teorías para explicar procesos mentales [416]. Posteriormente,

esa idea fue generalizada por el propio Newell en su propuesta de desarrollar teorías unificadas de las cognición y arquitecturas cognitivas [414]. También se mencionó que desde la propuesta de Newell, los investigadores de las ciencias cognitivas han hecho un enorme trabajo para explicar cómo la mente humana trabaja desarrollando arquitecturas cognitivas [276, 499, 286]. Todas esas investigaciones han sido formuladas en un marco denominado el paradigma computacional. El paradigma computacional ha sido elaborado por diferentes investigadores. Los primeros pasos en la formulación del marco fueron dados por Newell y Simon cuando enunciaron la hipótesis del sistema de símbolos físicos como requisito para que se genere el comportamiento inteligente [419]. Posteriormente, Newell propuso que los procesos cognitivos debían de describirse mediante una jerarquía de niveles como la que existe para describir las arquitecturas de los computadores, pero con la existencia del nivel de conocimiento en la cima de la jerarquía [413]. Independientemente de Newell, pero en la misma línea, David Marr propuso una jerarquía de niveles para explicar el sistema visual [341]. Marr propuso tres niveles para describir el procesamiento perceptivo de la información. Esta propuesta es conocida como la hipótesis trinivel en las ciencias cognitivas. Otro paso más en la formulación del paradigma computacional fue dado por Zenon Pylyshyn. Pylyshyn tomó la idea de una jerarquía de tres niveles de Marr, y la uso para proporcionar una fundamentación para las ciencias cognitivas [473]. La psicología cognitiva, la neurociencia computacional o la IA se han desarrollado en el marco del paradigma computacional. Sin embargo, el uso del paradigma computacional no está exento de críticas. Una de las más duras y famosas críticas es el argumento de la habitación china (AHC) de John Searle [518]. El argumento de Searle muestra que los procesos computacionales llevados a cabo por una máquina de Turing (o algoritmo) no puede explicar las características semánticas de la mente. Además, Hilary Putnam, uno de los creadores de la idea de representar la mente mediante una máquina de Turing y de la teoría funcionalista, decidió abandonar el funcionalismo debido a las dificultades que tiene la teoría en explicar el externalismo semántico [472].

En IA, el paradigma computacional es el marco donde toda la investigación se realiza. La conexión entre psicología y IA reside en que si la psicología puede describir un proceso en el paradigma computacional que ha generado un comportamiento, esa descripción puede ser usada por los investigadores de IA para crear un programa. Ese programa, al ser ejecutado en una computadora, hará que la computadora muestre el mismo comportamiento que la psicología observó cuando creó la descripción del proceso que lo generó. Una propuesta que se ha realizado en la IA para abordar las cuestiones semánticas en el paradigma computacional es la realizada por José Mira and Ana Delgado [393, 394, 391, 390]. Su propuesta es incorporar las ideas de Humberto Maturana y Francisco Varela al paradigma computacional, para hacer la diferenciación entre las semánticas internas y las semánticas del observador. La propuesta de Mira y Delgado formula el paradigma computacional como una jerarquía trinivel y dos dominios. Esos dos dominios son el dominio propio (DP) y el dominio del observador (DO).

La propuesta de Mira y Delgado se ha usado como una metodología para el diseño de sistemas artificiales que implementan exitosamente los métodos de resolución de problemas [391, 222, 221].

En la propuesta de Mira y Delgado un observador puede inyectar conocimiento a la descripción de un cálculo computacional para generar descripciones de DO pero la propuesta no profundiza sobre la naturaleza del DO. La propuesta de Mira-Delgado solo afirma que :

- Los seres humanos están en la categoría de observadores porque ellos pueden inyectar conocimientos usando lenguaje natural.
- Cada observador tiene un sistema de referencia asociado a él.

Ante esta situación, surgen varias cuestiones: ¿Cómo se puede diseñar un observador? ¿Qué significa exactamente inyectar conocimiento? ¿Cuánto conocimiento puede ser inyectado por un observador? ¿Qué es exactamente un sistema de referencia? ¿Qué hace el sistema de referencia de un observador?

Los trabajos pioneros de Newell, Marr, y Pylyshyn desarrollando el paradigma computacional han contribuido a desarrollar los fundamentos intelectuales de las ciencias cognitivas, incluida la IA, de un modo importante, y han abierto el camino a otras investigaciones [418, 3, 120]. El trabajo de Mira y Delgado, aunque menos conocido que los de Newell, Marr o Pylyshyn, muestra que la estructuración del paradigma computacional no estaba completa. Por lo tanto, dos cuestiones principales surgen en este contexto.

La primera cuestión es que si la propuesta de Mira y Delgado es capaz de describir cualquier valor de fasa. La segunda cuestión es si la limitación encontrada por Searle es una limitación del paradigma computacional o solo una limitación de las estructuraciones propuestas para el paradigma computacional hasta ahora. Así, en este capítulo, se analizan los límites de las distintas formulaciones del paradigma computacional. La investigación mostrará que la propuesta de Mira y Delgado, a pesar de ser la propuesta con más elementos de estructuración del paradigma computacional, tiene una limitación semántica que no le permite describir todos los valores de fasa. Por ese motivo, se propone una nueva estructuración del paradigma computacional que extiende a las anteriores mediante la incorporación de semánticas. Este nuevo formalismo, proporciona nuevas respuestas a cuestiones sobre el papel de la semántica en la generación de comportamiento. Estos resultados mostrarán que el paradigma computacional está lejos de estar agotado para explicar comportamientos y permite describir valores de fasa de sistemas exoactivos que son observadores.

Este capítulo está organizado de la siguiente manera. La sección 8.1 revisa las propuestas de Newell, Marr, Pylyshyn y Mira y Delgado. La sección 8.2 está dedicada a revisar conceptos matemáticos semánticos y fundamentos matemáticos de la teoría

de las categorías. La sección 8.3 estudia las limitaciones de la propuesta de Mira y Delgado. La sección 8.4 propone una solución a las limitaciones de la propuesta de Mira y Delgado, el marco-*S*. Finalmente, la última sección analiza los resultados que se han presentado en este capítulo.

## 8.1. Revisión de las propuestas para el paradigma computacional

Esta sección revisa las propuestas de Newell, Marr, Pylyshin y Mira y Delgado para estructurar el paradigma computacional. Las cuatro propuestas pueden ser vistas como una línea continua de investigación porque no se contradicen entre ellas, sino que cada propuesta es básicamente una extensión que completa a la anterior.

### 8.1.1. Jerarquía de niveles

El paradigma computacional tiene como uno de sus principales elementos la jerarquía de niveles. En el capítulo 6 ya se explicó el concepto de jerarquía de niveles que Newell propuso para describir y comprender los procesos cognitivos de máquinas y seres humanos. Tras la propuesta de Newell, Marr propuso analizar la percepción visual con estructura de tres niveles [341]. Los niveles que propuso fueron:

- Teoría Computacional. Este es el nivel más alto de la jerarquía. En este nivel se contestan las siguientes preguntas: ¿cuál es la meta de la computación? ¿cuál es la lógica por la que se puede implementar? y ¿por qué es apropiada?
- Representación y algoritmos. Este nivel está debajo del de teoría computacional. En este nivel se responden las siguientes preguntas ¿cómo se puede implementar la teoría computacional del nivel anterior? Concretamente, ¿cuál es la representación para los datos de entrada y de salida? y ¿cuál es el algoritmo para la transformar los datos de entrada en los de salida?
- Implementación física: Este nivel es el que está en el nivel inferior de la jerarquía. En este nivel se responde a la pregunta: ¿cómo puede realizarse físicamente la representación y el algoritmo?

Como se ha mencionado, la propuesta de Marr se conoce como la hipótesis trinivel. La figura de 8.1 representa gráficamente los niveles de la jerarquía propuesta por Marr.

El trabajo de Marr, aunque no se basara en la propuesta de Newell, puede verse cómo un análisis de la propuesta de Newell en el que la conclusión es que los niveles de la jerarquía pertenecen a tres clases de niveles distintas.

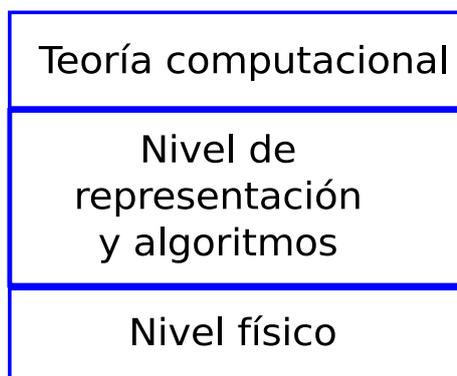


Figura 8.1: Jerarquía de niveles de Marr

Un paso más allá de la propuesta de Marr, fue dado por Zenon Pylyshyn. Pylyshyn propuso que los procesos mentales son procesos computacionales, pero no como una metáfora, sino como una hipótesis empírica seria [473]. En su propuesta, él usó la hipótesis trinivel de Marr para proponer el uso de una jerarquía también de tres niveles para describir los procesos cognitivos. Propuso los tres siguientes niveles: el nivel físico, el nivel de procesamiento simbólico y el nivel semántico. Pylyshyn afirma en su propuesta que el nivel de procesamiento simbólico y semántico se corresponden con los niveles simbólico y de conocimiento de Newell respectivamente [473]. La figura 8.2 muestra una representación gráfica de su jerarquía.



Figura 8.2: Jerarquía de niveles de Pylyshyn

Así, la propuesta de Pylyshyn lleva la hipótesis trinivel de Marr a un nuevo nivel, ya que pasa de que su dominio sea el de un proceso particular a ser el de todos los procesos cognitivos.

La propuesta de Mira y Delgado usa la propuesta trinivel[394]. Ellos mismos mencionan que su jerarquía de niveles es equivalente a los de Newell. Pero, aunque los niveles de la propuesta de Mira y Delgado se correspondan con las mismas cuestiones de Ne-

well y Pylyshyn, su propuesta incorpora una cuestión que la hace más una extensión de la de Newell. Esta extensión reside en la definición del nivel simbólico. Mira y Delgado proponen el concepto de *símbolo dinámico*, un concepto de símbolo más general que el de la propuesta de Newell [394, 392]. El concepto de símbolo dinámico permite que explícitamente la jerarquía de niveles se pueda aplicar a los sistemas conexionistas y el sistema nervioso. Así, ellos hacen la siguiente definición de símbolo en su propuesta:

*“ The symbols[in neurophysiological systems] are active and dynamic entities associated with specific patterns of spatio-temporal signals(electrical, chemical and electronics) that are presented repeatedly in a stable, independent and autonomous way and associated with specific references in the organism’s internal and external environment.”* [394] p.227

En su última formulación, Mira y Delgado usan las siguientes denominaciones para cada uno de los niveles: físico, simbólico e intencional.

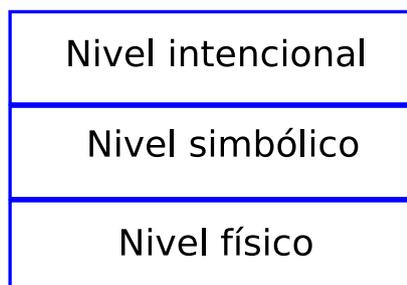


Figura 8.3: Representación gráfica de la jerarquía de niveles de Mira and Delgado [393]

Una cuestión que no se suele discutir en la bibliografía es porque la estructuración de paradigma computacional es importante. La razón es que la estructuración del paradigma computacional, no es sólo que permite la descripción de sistemas y su diseño, sino que determina cuánto se sabe sobre cómo se generan los comportamientos. Supóngase que hay dos sistemas que ante situaciones idénticas se comportan de manera distinta pero las descripciones que se tienen del proceso que genera los comportamientos son iguales. Entonces la estructuración es incompleta porque se debería de poder explicar donde reside la diferencia que provoca que se generen comportamientos distintos. También puede ocurrir a la inversa. Imagínese que se tienen dos sistemas que se comportan igual; pero se tienen dos descripciones diferentes, entonces la estructuración también es incompleta porque se debería poder decir porque están generando el mismo comportamiento.

Usando la terminología de Mira y Delgado [393], el nivel intencional es un nivel muy abstracto, ya que está situado sobre el nivel simbólico. Aun así, es valioso para explicar

procesos de generación de comportamientos que con los otros niveles no se puede apreciar que son iguales. La necesidad de los niveles físico y simbólico para comparar procesos computacionales es intuitiva; pero el propósito del nivel de conocimiento puede parecer más opaco. Sin embargo, la jerarquía de niveles es una estructuración importante del paradigma computacional. Si se tienen dos sistemas cuyo comportamiento es el mismo, existen las siguientes posibilidades. Los dos sistemas tienen la misma descripción en los tres niveles. La siguiente es que sus cálculos son diferentes a nivel físico, pero si tienen los mismos símbolos, entonces los cálculos serán los mismos a nivel simbólico y a nivel intencional. La última opción es que, si los sistemas son físicamente diferentes y además tienen diferentes símbolos, entonces los cálculos de ambos tienen que ser iguales en el nivel intencional.

### 8.1.2. Dominios

A pesar de la capacidad de las jerarquías trinivel para describir procesos, hay un problema con esa estructuración del paradigma computacional. El problema es que la jerarquía de lenguajes puede solo describir estados computacionales y operaciones que el sistema realiza, pero la semántica no es descrita en los estados u operaciones de un mecanismo computacional. Pylyshyn en su propuesta era consciente de este problema e hizo la siguiente especulación sobre la mente humana y cómo los seres humanos podrían proporcionar semántica a las máquinas :

*“This is a different kind of designation because it is not wired into the machine (for which all symbols tokens are meaningless, except that certain ones can be made to cause the primitive of primitive operations). The designation is provided by a person who takes the symbols to be about something - that is, the person gives the symbols a semantic representation .”* [473]

Algunos años antes de la especulación de Pylyshyn, Humberto Maturana y Francisco Varela se enfrentaron al mismo problema, aunque en un contexto ligeramente diferente, el del sistema nervioso. Su propuesta como solución al problema se basó en el concepto de observador [347, 590, 346]. Maturana y Valera propusieron la existencia de dos dominios que están conectadas por los observadores y su sistema de referencia. Para explicar su propuesta de observador y dominio, ellos propusieron el siguiente ejemplo:

*“Imagine a person who has always lived in a submarine. He has never left it and has been trained how to handle it. We are now standing on the shore and see the submarine slowly gracefully surfacing. We then get on the radio and tell the navigator inside: ‘Congratulations! You avoided the reefs and surfaced beautifully. You really know how to handle the submarine’. The navigator in the submarine, however, is perplexed:*

*”What’s all this about reefs and surfacing? All I did was push some a lever and turn knobs and make certain relationships between indicators as I operated the levers and knobs. It was all done in a prescribed sequence which I am used to. I didn’t do any special maneuver, and on the top of that, you talk to me about a submarine. You must be kidding!”*[346]p.136-137.

La persona que está en la costa es el observador porque conecta ambos dominios: el interno y el externo. La propuesta de Maturana y Varela fue adoptada por Mira y Delgado para estructurar el paradigma computacional. Así, Mira y Delgado propusieron la introducción de dos dominios en la estructuración del paradigma computacional [394, 392, 390]: el dominio propio(DP) y el dominio del observador(DO). Esto implica que en cada uno de los niveles hay dos dominios, y que a cada descripción de un proceso computacional se le asignaría un nivel y un dominio. La figura 8.4 representa gráficamente la propuesta de Mira y Delgado.

OWN DOMAIN	OBSERVER DOMAIN
Nivel Intencional	Nivel Intencional
Nivel simbólico	Nivel simbólico
Nivel físico	Nivel físico

Figura 8.4: Representación de la propuesta de niveles y dominios de Mira y Delgado

El DP contiene la descripción operacional de un cálculo. Mira y Delgado lo explican de la siguiente manera::

“[En el DP] *There are only data structures and formal algorithms of their own*” [391]p.17

Mira y Delgado hacen la siguiente afirmación sobre el DP:

*“The algorithm and data structure of the program being executed at that moment cannot be deduced from the detailed knowledge of what is happening in the computers’s inverters and registers.”* [394]p.217.

Para entender la anterior afirmación, se puede usar el siguiente ejemplo. Supóngase que se tienen dos sistemas expertos con el mismo algoritmo para la toma de decisiones,

pero cada uno toma decisiones para un dominio distinto. Uno sistema experto contrala la temperatura de un alto horno. El otro sistema experto toma decisiones para mantener la distancia de un satélite a la tierra. Ambos sistemas expertos reciben los datos de entrada por consola y muestra los datos de salida por pantalla. Si se observa el proceso computacional de uno de los sistemas expertos que solo contiene números y operaciones matemática, no se puede saber si se está observando el proceso del sistema que toma decisiones sobre el alto horno o el del satelite.

Respecto al DO, la propuesta es que el DO contiene la descripción del DP. Así, la descripción del DO permitiría a un observador entender cómo el sistema está relacionado con su entorno. Esa descripción es creada por el observador. Mira y Delgado propusieron los dos dominios con la intención de introducir el concepto de un observador externo y un sistema de referencia en el paradigma computacional [390]. Respecto al sistema de referencia, Mira y Delgado afirman que:

*“[los sistemas de referencia] in which entities, relations, and their meanings are re-presented.[394]p.218.*

Así, cada observador estaría acoplado a un sistema de referencia que le permitiría crear las descripciones del DO gracias a que el sistema de referencia contendría las entidades, la relaciones entre ellas y sus significados.

Ahora bien, ¿qué aporta tener los dos dominios al paradigma computacional? La respuesta es que el DO ayuda a describir correctamente que observadores tengan comportamientos distintos en determinados casos que no se podría sin el uso de los dominios. Esto se puede ver con un sencillo ejemplo. Imagínese que una computadora muestra un mensaje en la pantalla y permanece esperando a recibir una respuesta desde la consola. El mensaje dice que mire la temperatura que marca el termómetro de la habitación y lo introduzca por teclado. Si el usuario que está en la habitación no entiende la lengua en la que está escrito el mensaje entonces el usuario no puede escribir una respuesta adecuada. Sin embargo, si el usuario entiende la lengua, entonces el usuario introducirá la temperatura que marca el termómetro y la computadora podrá tomar una decisión sobre si encender o no el aire acondicionado. La diferencia entre los dos usuarios estaría en que ellos tienen un sistema de referencia diferente, y, por lo tanto, una descripción diferente en el DO. El DO de uno de los usuarios tiene una semántica para el mensaje de la pantalla mientras que le otro carece de ella. Así, uno de los observadores puede inyectar conocimiento a la computadora mientras que el otro observador no es capaz de hacerlo. Por tanto, el DO ayuda a diferenciar situaciones donde unos observadores pueden inyectar conocimiento y otros no.

En base a todo lo anterior, la propuesta de Mira y Delgado es la que aporta la estructuración más completa del paradigma computacional porque, aunque todas las propuestas tienen una descripción multinivel que son equivalentes entre ellas, añade el

concepto de dominios que no colisiona con el concepto de jerarquía ya que se superpone y permite hacer distinciones en las descripciones que no se pueden hacer sin los dominios.

## 8.2. Fundamentos matemáticos

Para abordar con precisión la estructuración del paradigma computacional se han usado diferentes formalismos matemáticos en la investigación que se ha llevado a cabo. Esta sección realiza una breve introducción a esos formalismos que se han usado. El primer apartado de esta sección está dedicado a las distintas nociones matemáticas semánticas y sus formalismos. El segundo apartado de esta sección hace una introducción a la teoría de las categorías que será usada en la nueva propuesta para reestructurar el paradigma computacional.

### 8.2.1. Semántica

La semántica es un elemento importante en las propuestas y discusiones del paradigma computacional que se han hecho. Sin embargo, en esas propuestas o discusiones no se da una definición formal de lo que los autores quieren decir cuando hablan de semántica. En matemáticas e informática teórica, existen diferentes nociones semánticas con sus definiciones formales. Este apartado va a revisar los conceptos matemáticos que existen para abordar la semántica que permiten analizar el concepto de semántica que existe en las distintas propuestas para estructurar el paradigma computacional.

La investigación matemática de la semántica ayudó al desarrollo de la lógica matemática y la informática en el siglo XX. Las investigaciones han conducido al desarrollo de dos nociones semánticas diferentes. La primera noción semántica fue desarrollada por Alfred Tarski[561] en sus trabajos para definir formalmente el concepto de verdad en las matemáticas. A este tipo de semántica, se la denominará en este texto *semántica-t*. Otra noción formal de semántica diferente se ha desarrollado en informática para dar el significado de las instrucciones de los lenguajes de programación respecto a los efectos de su ejecución por parte de una computadora[624]. A este otro tipo de semántica se le denominará, *e-semántica*.

#### 8.2.1.1. Semántica-e

En informática existen diferentes métodos para abordar el problema de cómo dar el significado de las instrucciones de un lenguaje de programación. Por ejemplo, algunos de esos métodos son: las máquinas abstractas, las semánticas operacionales estructurales de paso pequeño, las semánticas operacionales estructurales de paso grande, las

semánticas denotacionales o las semánticas axiomáticas. Cada método de e-semántica es desarrollado para dar el significado de expresiones matemáticas y comandos en lenguajes de programación. Las semánticas operacionales estructurales es uno de los métodos más usado en informática. Este método describe un lenguaje de programación mediante axiomas y reglas que especifica cómo las expresiones del lenguaje son evaluadas. A continuación, se presenta un pequeño ejemplo de un pequeño lenguaje imperativo, SIL. La gramática de SIL es:

$$\begin{aligned}
 P & ::= C \\
 C & ::= \text{skip} \mid C;C \mid l := E \mid \text{if } B \text{ then } C \text{ else } C \mid \text{while } B \text{ do } C \\
 E & ::= !l \mid n \mid E \text{ op } E \\
 \text{op} & ::= + \mid - \mid * \mid / \\
 B & ::= \text{True} \mid \text{False} \mid E \text{ bop } E \mid \neg B \mid B \wedge B \\
 \text{bop} & ::= > \mid < \mid =
 \end{aligned}$$

$P$  es el símbolo para programa,  $C$  para comando,  $E$  para expresión de enteros y  $B$  para booleano. Para distinguir el contenido de una variable de la propia variable, cuando se use el contenido de una variable se escribirá  $!l$  para la variable  $l$ . La semántica operacional estructural proporciona el significado de un programa escrito en SIP a través de una definición inductiva del proceso de evaluación usando axiomas y reglas. Se escribe  $(E, s) \Downarrow n$  para indicar que el resultado de evaluar la expresión  $E$  en el estado  $s$  es  $n$ . El significado de las expresiones de enteros, el comando saltar *skip*, y el comando asignación  $:=$  se formalizan a continuación.

El significado de una expresión de enteros es dado por axiomas y la regla de evaluación. Los axiomas son:

- $\overline{(n, s) \Downarrow (n, s)}$  para todos los enteros.
- $\overline{(!l, s) \Downarrow (n, s)}$  if  $l \in \text{dom}(s)$  y  $s(l) = n$ .

y la regla es

$$\frac{(E_1, s) \Downarrow (n_1, s') \quad (E_2, s') \Downarrow (n_2, s'')}{(E_1 \text{ op } E_2, s) \Downarrow (n, s'') \text{ if } n = n_1 \text{ op } n_2}$$

El significado del comando *skip* viene dado por el siguiente axioma:

$$\overline{(\text{skip}, s) \Downarrow (\text{skip}, s)}$$

El significado del comando  $:=$  viene dado por la regla:

$$\frac{(E, s) \Downarrow (n, s')}{(l := E, s) \Downarrow (skip, s[l \mapsto n])}$$

donde  $s[l \mapsto n]$  denota la función  $s'$  que coincide con  $s$  excepto en que el valor  $n$  está asociado a  $l$ .

### 8.2.1.2. Semántica- $t$

Una semántica- $t$  se define mediante tres elementos: un modelo, un lenguaje formal y una interpretación. La interpretación conecta el lenguaje formal con el modelo. El significado de una fórmula del lenguaje es la parte del modelo que asigna la interpretación. Tarski usó el siguiente ejemplo para explicar la idea principal que subyace en su trabajo:

*“La nieve es blanca” es verdad si y solo si la nieve es blanca*

La expresión tiene dos objetos que pertenecen a clases diferentes: palabras y objetos del mundo físico. Las palabras están señaladas por las comillas. El objeto físico está sin comillas. Esos dos objetos están relacionados por la expresión *si y solo si*. Una estructura con esos tres elementos con la forma, “ $X$ ” es verdad si y solo si  $\mathbf{X}$ , se denomina esquema-T. Una interpretación es construida mediante esquemas-T. A continuación, para ilustrar formalmente los conceptos se muestra un ejemplo:

**Definición 6** (Modelo). *Un modelo es una estructura matemática. Un tipo de estructura que se usa por su generalidad es una tupla que contiene tres conjuntos: un conjunto de elementos (denominado universo), un conjunto de funciones cuyos dominios y codominios es el universo del modelo y un conjunto de relaciones definidas sobre el universo. Así, en una notación general, un modelo se define como:*

$$\mathcal{A} = \langle \mathbf{A}, \langle f_i \rangle_{i \in I}, \langle R_j \rangle_{j \in J} \rangle$$

y se dice que el modelo es de tipo  $\langle \mu, \delta \rangle$

**Definición 7** (Lenguaje formal). *Un lenguaje formal es un conjunto de fórmulas,  $L = \{\varphi, \dots\}$ . Cada fórmula es una combinación de elementos que pertenecen a un conjunto denominado alfabeto. Un alfabeto tiene diferentes tipos de elementos: un conjunto de variables  $\mathcal{V}$ , símbolos lógicos, funtores  $\langle f_i \rangle_{i \in I}$ , relatores  $\langle R_j \rangle_{j \in J}$  y símbolos auxiliares. Las combinaciones de los símbolos del alfabeto que forman parte del lenguaje vienen determinadas por un conjunto de reglas.*

**Definición 8** (Asignación). *Una asignación es un elemento que determina el valor de cada variable asignándole un elemento concreto del universo del modelo.*

$$\mathcal{I} : \mathcal{V} \rightarrow \mathbf{A}$$

**Definición 9** (Interpretación ). *Una interpretación da una definición de verdad inductiva y recursiva para las fórmulas de un lenguaje respecto a un modelo mediante esquemas- $T$ . La recursividad permite que la verdad de las fórmulas complejas quede definida a partir de las formulas simples que las componen. Una interpretación determina cuando una estructura matemática  $\mathcal{A}$  y una asignación  $\mathcal{I}$  son un modelo de una formula  $\varphi$ , o, dicho de otra manera, si  $\varphi$  es verdad en la estructura  $\mathcal{A}$  con la asignación  $\mathcal{I}$ , esto se denota con  $\mathcal{AI} \models \varphi$ . Los detalles de la definición inductiva concreta para un lenguaje lógico puede encontrarse en múltiples textos universitarios [336, 335].*

Un ejemplo de uso de la notación anterior es el siguiente. Dada la estructura de los números naturales con la relación de orden  $(\mathbb{N}, <)$  y la formula  $\varphi \equiv \forall x \exists y \ x < y$  se puede decir que:

$$(\mathbb{N}, <) \models \varphi$$

Las semánticas-t han sido intensamente investigadas durante el siglo XX. Aquí, solo se han mencionado estructuras con un único universo, pero se han investigado estructuras más complejas, generalizando la idea de un universo y permitiendo que la estructura tenga más de un dominio [335]. También se han estudiado generalizaciones del concepto de interpretación con la propuesta de estructuras de muchos mundos de Kripke permitiendo que la estructura cambie mediante relaciones de accesibilidad.

### 8.2.2. Teoría de la categoría

La noción de categoría fue definida en 1945 [142]. La teoría de las categorías se ha desarrollado hasta llegar a ser una rama muy importante de las matemáticas puras. La teoría permite una gran flexibilidad describiendo objetos matemáticos debido a que es el marco matemático más abstracto que existe.

Una categoría  $\mathbf{C}$  consiste de una colección de *objetos*  $obj_{\mathbf{C}}$  y una colección de *morfismos*  $Mor_{\mathbf{C}}$ . Los objetos son denotados por letras mayúsculas:

$$A, B, C, D, \dots, X, Y, \dots$$

y los morfismos son denotados por letras minúsculas:

$$f, g, h, \dots$$

Una categoría consta formalmente de las siguientes características:

- Un morfismo tiene asignados dos elementos. Uno es denominado dominio y el otro codominio, y ambos son elementos de  $obj_{\mathbf{C}}$ . Cuando el dominio de  $f$  es  $A$  y el codominio de  $f$  es  $B$  se escribe  $f : A \rightarrow B$
- Si el morfismo  $f : A \rightarrow B$  y  $g : B \rightarrow C$  están en  $Mor_{\mathbf{C}}$ , hay un morfismo en  $Mor_{\mathbf{C}}$ , denominado morfismo composición, que tiene la forma  $g \circ f : A \rightarrow C$ .

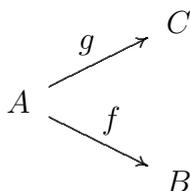


Diagrama 8.2.1: Diagrama de una categoría.

- Para cada objeto  $A$  de  $obj_{\mathbf{C}}$ , existe un morfismo en  $Mor_{\mathbf{C}}$ , denominado morfismo identidad, que tiene la siguiente definición  $1_A : A \rightarrow A$
- Para todo morfismo  $f : A \rightarrow B$ ,  $g : B \rightarrow C$  y  $h : C \rightarrow D$  (con  $A, B, C$  y  $D$  no necesariamente diferentes) de  $Mor_{\mathbf{C}}$  se cumple que:

$$h \circ (g \circ f) = (h \circ g) \circ f$$

- Para cualquier morfismo  $f : A \rightarrow B$  se cumple que,

$$id_B \circ f = f \text{ and } f \circ id_A = f$$

La teoría de las categorías a menudo usa un método gráfico para representar sus afirmaciones o deducciones porque las afirmaciones sobre objetos y morfismos se vuelven complicados rápidamente. La representación gráfica es denominada, diagrama. Un diagrama de una categoría  $\mathbf{C}$  es una colección de vértices y aristas dirigidas, etiquetados consistentemente con objetos y morfismos de la categoría en cuestión. A las aristas dirigidas del diagrama se las denomina *flechas*. Por ejemplo, siendo una categoría  $\mathbf{C}$  con los objetos  $A, B$  y  $C$ , y los morfismos  $f : A \rightarrow B$  y  $g : A \rightarrow C$ , su diagrama es el del diagrama 8.2.1.

Este diagrama se dice que *conmuta*, o que es *commutativo* porque cumple que cada par de vértices  $X$  e  $Y$ , todos los caminos entre  $X$  e  $Y$  determinan el mismo morfismo de la categoría  $\mathbf{C}$ . Se puede ver fácilmente que el operador composición determina un camino entre dos vértices. El diagrama 8.2.2 es conmutativo.

por lo tanto, se cumple que:

$$h \circ f = j \circ g$$

El diagrama de una categoría también puede resultar no conmutativo. El diagrama 8.2.3 es el de una categoría cuyo diagrama es no conmutativo.

por lo tanto, se cumple que:

$$h \circ f \neq j \circ g$$

En teoría de las categorías una propiedad se define muchas veces mediante un diagrama conmutativo de la categoría. Los diagramas conmutativos y no conmutativos

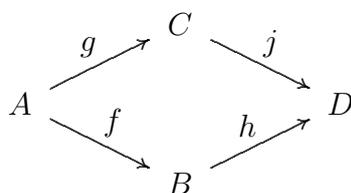


Diagrama 8.2.2: Diagrama conmutativo.

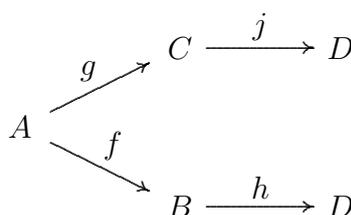


Diagrama 8.2.3: Diagrama no conmutativo.

serán un elemento importante para los razonamientos y los resultados obtenidos en la investigación que se presenta en este capítulo.

Otro concepto importante de la teoría de las categorías es el concepto de *functor*<sup>1</sup>. La definición de functor se realiza a partir de dos categorías. Un functor  $F$  de la categoría  $\mathbf{C}$  a la categoría  $\mathbf{C}'$  es una función que asigna elementos según las siguientes reglas:

- Para cada objeto  $A$  de  $obj_{\mathbf{C}}$  el functor le asigna un objeto de  $obj_{\mathbf{C}'}$ ;
- Para cada  $f : A \rightarrow B$  de  $Mor_{\mathbf{C}}$  el functor le asigna un morfismo  $F(f) : F(A) \rightarrow F(B)$  de  $Mor_{\mathbf{C}'}$ , que cumple que:
  1.  $F(id_A) = id_{F(A)}$ , para todos los objetos  $A$  de  $obj_{\mathbf{C}}$ .
  2.  $F(g \circ f) = F(g) \circ F(f)$ , si  $g \circ f$  está definido.

Un functor es denotado por  $F : \mathbf{C} \rightarrow \mathbf{C}'$ .

El lector tiene que tener en cuenta que una categoría puede además tener categorías como objetos y funtores como morfismos. Existen múltiples textos universitarios que pueden ser consultados para más detalles sobre ella [603, 285].

<sup>1</sup>La palabra functor aquí tiene un significado diferente de en teoría de modelos en la que se usa para un símbolo del lenguaje que denota una función en la estructura.

### 8.3. Limitación de la propuesta de Mira y Delgado

La meta principal de la IA es lograr la reproducción de todas las habilidades cognitivas que poseen los seres humanos. Como se ha mostrado en la sección 8.1 la propuesta de Mira y Delgado es la más completa para estructurar el paradigma computacional [393]. La propuesta de Mira y Delgado se ha usado efectivamente en el diseño de sistemas informáticos, pero la cuestión que se busca abordar aquí es si la estructuración del paradigma computacional de Mira-Delgado es capaz de describir cualquier valor de fasa.

La estructuración de Mira y Delgado ha sido propuesta para modelar e interpretar la función neuronal de los seres humanos y realizar “*neurofisiología inversa*” [394, 392]. Esa propuesta implica que la estructuración de Mira-Delgado podría ser adecuada para describir cualquier valor de la fasa. Además, debería ser capaz de describir el diseño de un sistema computacional que puede observar e inyectar conocimientos porque la propuesta de Mira-Delgado considera a los seres humanos observadores, y, por ende, capaces de hacer esas acciones. Sin embargo, no está claro si la estructuración de Mira-Delgado puede interpretar toda la actividad neuronal. La propuesta de Mira y Delgado es interesante para el objetivo de este capítulo porque si se encuentra alguna actividad neuronal que nos puede ser descrita con la estructuración de Mira y Delgado, eso significará que no es un marco lo suficientemente expresivo para describir todos los valores de fasa. El observador y la capacidad para inyectar puede ser asimilado dentro de los conceptos que maneja la filosofía de la mente y la neurociencia con la capacidad de lectura de mente (*mindreading*). La lectura de mente es la capacidad que poseen los seres humanos para atribuir pensamientos e intenciones a otras personas. El proceso de lectura de mente ha sido ampliamente investigado y discutido por la filosofía de la mente y la ciencia cognitiva. El revisar el proceso de lectura de mente está fuera de los objetivos de este capítulo. Sin embargo, hay un hecho importante que debe ser tomado en cuenta de la investigación de esas disciplinas: cuando un ser humano está llevando a cabo un proceso de lectura de mente, se está produciendo una actividad neuronal en su cerebro humano que no ocurre cuando el ser humano no está lo está llevando a cabo [592, 593]. Por lo tanto, la neurociencia prueba que la lectura de mente es necesariamente un proceso neuronal. La propuesta de Mira y Delgado afirma que en el DP no hay otra semántica que la operacional. Considerando esos dos hechos, se podría considerar que el proceso de lectura de mente no es un cálculo. Sin embargo, una de las premisas de la premisa de Mira y Delgado es la siguiente:

“*they[the circuits of nervous system] are the physiological support of a calculus*” [394]p.212.

Así, la propuesta de Mira y Delgado afirma que toda actividad neuronal está en el DP porque solo tiene semántica operacional. Sin embargo, por las investigaciones en

neurociencia se sabe que los procesos de lectura de mente, que se llevan a cabo mediante actividad neuronal, son una descripción con conocimiento inyectado. Así, el proceso de lectura de mente debería estar en el DO. Por lo tanto, se ha encontrado una contradicción porque es imposible que una descripción pertenezca al mismo tiempo al DO y al DP. La conclusión es que la propuesta de Mira y Delgado tiene límites a la hora de describir procesos, y los procesos de lectura de mente quedan fuera de sus límites

Los procesos de lectura de mente han mostrado que la estructuración de Mira y Delgado tiene límites, pero la contradicción proporciona la imposibilidad de que todos los procesos computacionales en el DP solo tengan semántica operacional. Muy probablemente Mira y Delgado eran conscientes de las limitaciones de su propuesta. Mientras que en la formulación ellos colocaban las semánticas no operacionales solo en el DO, en las conclusiones de su propuesta de usar su estructuración del paradigma computacional para interpretar el cálculo que lleva a cabo el sistema nervioso, apuntan lo siguiente <sup>2</sup>:

*“it is a concept[semantics] inherent in the EOD and in the own domain it emerges as a result of the sensitivity of the nervous tissue that enables this kind of organization.”* [394]p.236

Además, en su último artículo ellos proponen la existencia de un caso especial con el autoobservador para abordar la conciencia.

*“We include as a singular case the figure of the self-observer for the distinction between experience and verbal communication. In other words, between conscious individual experience of the perception phenomenon, which cannot exist outside the language of each self-observer, and what is communicable to other external observers sharing the same natural language.”* [393]p.794.

Sin embargo, en el artículo no se hace referencia a las propiedades del dominio de ese observador. Así, todo apunta a que conocían perfectamente que de algún modo el DP debe tener alguna semántica en el caso del sistema nervioso, ya que, sin una, el sistema no podría ser explicado completamente de una manera científica. Este resultado conduce a que es necesario considerar una estructuración diferente del paradigma computacional para describir todos los valores de fase que posee la naturaleza.

Para proponer una estructuración alternativa del paradigma computacional, es necesario analizar cuál es el problema que se ha descubierto sobre los procesos de lectura de mente. Antes de llevar a cabo el análisis, conviene hacer una puntualización sobre la

---

<sup>2</sup>La terminología de la propuesta de Mira y Delgado en inglés para los dominios es *external observer domain*(EOD) para el dominio del observador y *own domain* (EO) para el dominio propio.

terminología de Mira y Delgado. Usan la expresión descripción operacional de un modo genérico, pero como se ha mencionado, las semánticas operacionales son un formalismo muy concreto. Así que, a partir de ahora, se sustituirá la expresión descripción operacional por la expresión semánticas-*e* (semánticas de ejecución) para la descripción de procesos computacionales.

El punto de partida del análisis es la afirmación de Mira y Delgado de que no hay semántica en el DP. Usando la terminología que se ha revisado en la sección anterior, Mira y Delgado lo que afirman es que solo hay semánticas-*e* en el DO y no tiene ninguna semántica-*t*. Si esa afirmación fuera verdad, estaría llevando a cabo sus procesos computacionales sin otro tipo de semántica que semántica-*e*. Cuando una computadora comercial (ej. PC, portátil, etc.) realiza un cálculo cuya entrada de datos es controlada por consola o cargados desde un archivo, y la salida es por pantalla o almacenando los datos en un archivo, el sistema solo tiene una semántica-*e*. Por lo tanto, la propuesta de Mira y Delgado es correcta cuando es usada para describir DO de computadoras comerciales. Sin embargo, ¿se puede generalizar ese hecho y afirmar que todos los sistemas computacionales poseen solo semántica-*e* en el DO? Mira y Delgado asumieron que el criterio de Varela era válido para separar el DO del DP para todos los sistemas computacionales:

*“Does a given entity X play a causal role in the biological system? If yes, then X belongs to OD, if not then X belongs to EOD.”* [394]p.220.

Así es importante analizar el criterio escogido por Mira y Delgado. Su traducción a una terminología más técnica está diciendo que en el dominio propio hay semántica-*e* pero no semántica-*t*. Para ese análisis, se ha usado el caso de la computadora que pregunta sobre la temperatura de la habitación en el apartado 8.1.2. En ese ejemplo, la computadora requiere un observador que posea una semántica-*t* para los símbolos de su mensaje, para que el observador pueda inyectar conocimiento que requiere para poder realizar la decisión correcta. Si el observador no posee una semántica-*t* para los símbolos del mensaje, entonces el mensaje no puede ser contestado. Así, el observador humano puede ser substituido por otra computadora que tenga sensores de temperatura. Así, la computadora primera envía un mensaje a la nueva computadora y la nueva computadora responde enviándole la temperatura a la que está la habitación. Observando el ejemplo objetivamente, se tiene que concluir que la nueva computadora tiene una semántica-*t* para los símbolos y puede poseer conocimiento que la primera computadora necesita para tomar una decisión y poder realizar correctamente su tarea.

El lector podría preguntarse a sí mismo: ¿cómo es posible que la nueva computadora tenga una semántica-*t* ? Para responder a esa pregunta, hay que ver que ocurre en el nivel físico. Mira y Delgado hicieron la siguiente afirmación:

*“In the physical level’s OD, processes cannot be separated from the processors reali-*

*zing them.*" [394]p.219

En el ejemplo de los ordenadores, hay un elemento en el que hay que centrarse: los sensores. Un sensor es un sistema computacional que posee transductores. Un transductor convierte una señal de una forma de energía en una señal de otro tipo de energía. Desde el punto de vista funcional, esa conversión debe ser considerada una computación porque el transductor genera una salida que depende de una entrada. La clave de la computación realizada por el transductor es que el dominio y el codominio de la computación son diferentes. Por lo tanto, hay una conexión entre el estado físico que está produciendo la entrada y la señal que se genera como salida. Esto es exactamente una semántica-*t* porque el transductor genera un valor numérico particular solo si la temperatura de la habitación es una concreta (considerando que el sensor está calibrado y funcionando). Por lo tanto, una semántica-*t* describe un proceso computacional y un elemento que describe un sistema computacional tiene un significado si este ha sido generado mediante una semántica-*t*. Así, hay sistemas computacionales que tienen una semántica-*e* y una semántica-*t*. Las afirmaciones de Mira y Delgado así como las de Maturana y Varela concernientes a la existencia únicamente de semántica-*e* en el DP no puede ser aplicada a todos los sistemas porque hay sistemas que tienen semántica-*t*.

Para contrastar más la conclusión a la que se ha llegado se ha analizado otro caso más. En esta ocasión partiendo del caso de los dos sistemas expertos también del apartado 8.1.2. En ese caso, se tenían dos sistemas expertos con el mismo algoritmo cuyas entradas de datos era por consola y la salida por pantalla. Uno tomaba decisiones para el control de la temperatura de un alto horno y el otro para la distancia de un satélite a la tierra. Ahora, se van a tener los mismos sistemas expertos pero esta vez con la diferencia de que la entrada de datos será mediante sensores que miden la temperatura en un caso y la distancia en el otro. Así, ahora, aunque las reglas son las mismas, su disparo depende del estado físico del entorno. Ahora, cuando los cálculos de ambos sistemas computacionales son examinados en el nivel físico, sí pueden ser distinguidos porque los transductores de los sensores son diferentes. Contrastando el primer caso en el que los sistemas tienen entrada por consola y este nuevo caso en el que tienen entrada por sensores, se muestra que hay sistemas que tienen y no tienen semántica-*t*

Por lo tanto, la conclusión es que un sistema computacional puede tener semántica-*e* y semántica-*t* en el dominio propio. Este hecho nunca fue considerado por Varela. Una cuestión que puede surgir ante la situación que se ha encontrado es: ¿por qué si dos sistemas computacionales tienen el mismo poder computacional tienen características distintas. Ante esa pregunta, lo primero a tener en cuenta es que hay diferentes modelos computacionales y no todos los modelos computacionales pueden resolver los mismos problemas computacionales debido a las características particulares de los problemas [138]. Pero hay otro hecho más importante a tener en cuenta, el poder computacio-

nal de un sistema hace referencia a una característica muy específica. Concretamente, el poder computacional hace referencia al límite fijado por una máquina de Turing computando funciones entre  $\mathbb{N}$ . Una semántica- $t$  es una característica diferente al poder computacional, porque una semántica- $t$  es una función cuyo dominio es un conjunto de estados de un entorno a un conjunto de símbolos. Es importante notar que esa última afirmación sobre que se tratan de funciones de distinto tipo no implica que una semántica- $t$  no es una función computacional. La máquina de Turing pertenece a un enfoque de la computación que se centra en un mecanismo muy concreto. Otro enfoque de computación es el enfoque funcional. Como ha sido mencionado anteriormente, una semántica- $t$  es una función computacional (como la explicación del sensor demuestra). Por lo tanto, debería ser considerado un elemento para realizar una descripción computacional.

Hasta ahora se han presentado casos que muestran que en el DP no hay solo semántica- $e$ . A continuación, se van a abordar los razonamientos que Maturana y Varela emplearon para afirmar que en el DP no existen semántica- $t$ , ya que la estructuración del paradigma computacional de Mira y Delgado se basaba en su propuesta [394]. Maturana y Varela basaron su propuesta en el experimento de Roger Sperry en el cual se rotaba la retina de una rana 180 grados [542]. En el experimento, se vio que ranas en las que sus retinas habían sido rotadas 180 grados cuando lanzaban su lengua para alimentarse de moscas, la trayectoria de su lengua estaba alterada el mismo número de grados que la retina había sido rotada respecto a la línea recta entre la boca de la rana y la mosca. Maturana y Varela analizan el experimento para concluir que no hay semántica en su DP [346]. Ellos señalan primero el hecho de que la rana no corrige su comportamiento mientras que el observador es consciente de que su comportamiento está alterado. En base a ese hecho, argumentan que el observador tiene una semántica que le permite conocer que el comportamiento de la rana es patológico, y la rana carece de esa semántica ya que no es consciente de que su comportamiento es patológico. Sin embargo, hay un problema en ese análisis. El problema es que ellos omiten una cuestión importante, *¿por qué hace la rana lo que hace?* La rana lanza su lengua porque la imagen del insecto en la retina marca comida para la rana en una posición específica [4]. Aunque Maturana y Varela tienen razón cuando dicen que hay una correlación entre las coordenadas en la retina y las contracciones musculares que mueven la lengua; no están tomando en consideración el hecho de que hay una correlación entre los estados físicos del entorno y las coordenadas de la retina, es decir, una semántica- $t$ . La operación de rotar la retina 180 grados también modifica la semántica- $t$ . Esto es porque la rana lanza su lengua en la dirección equivocada. A lo largo del tiempo, la evolución ha conformado la compatibilidad de la semántica- $t$  con la semántica- $e$  de la fisiología de las ranas. En el experimento, cuando se realiza la rotación de la retina en la rana, su semántica- $t$  es remplazada por otra, pero la nueva semántica- $t$  es incompatible con la semántica- $e$  del sistema nervioso de la rana. Así, lo que está produciendo el comportamiento patológico de la rana desde el punto de vista computacional es la

incompatibilidad entre la semántica-*t* y la semántica-*e*. Contrariamente a lo que Maturana y Varela afirmaron, y Mira y Delgado asumieron, el experimento de Sperry no muestra que solo hay semántica-*e* en el DP; Lo que el experimento muestra es lo que ocurre cuando la semántica-*t* y la semántica-*e* son incompatibles. Por lo tanto, la conclusión a partir del experimento de Sperry de que la rana solo tiene semántica-*e* (semántica operacional) es incorrecta. La conclusión tras la discusión que se ha llevado a cabo se puede enunciar de la siguiente manera:

**Una semántica es una función computacional y puede estar en el dominio propio de un sistema computacional.**

Así, la estructuración del paradigma computacional asumiendo dominios debe permitir la descripción de semántica-*t* en el DP. En base al resultado obtenido, se puede afirmar que al menos hay dos tipos de sistemas computacionales en función de la semántica que tienen en su DP. A estos dos tipos de sistemas se les denominará: libres y ligados. Un sistema de semántica-*t* libre realiza un proceso computacional sin que tenga unión que lo ligue a los estados físicos de su entorno. Un sistema de semántica-*t* ligada es un sistema computacional que hace un proceso computacional con una semántica-*t* porque sus entradas están ligadas a estados físicos específicos del entorno.

La estructuración del paradigma computacional de Mira y Delgado es válida cuando los sistemas que se quieren describir en el DP son sistemas de semántica-*t* libre, por eso ha sido efectiva en las investigaciones en la que se ha aplicado [391, 222, 221]. Pero no es válida cuando es un sistema con semántica-*t* ligada, y por lo tanto para describir todos los valores de la fasa hace falta una estructuración más compleja del paradigma computacional de la propuesta por Mira y Delgado. Así, en la investigación se ha abordado el desarrollo de métodos para la descripción de la semántica-*t* de sistemas computacionales.

## 8.4. Descripción de la semántica-*t* de sistemas computacionales

La investigación que se ha llevado a cabo para lograr métodos que permitan describir la semántica-*t* de un sistema computacional se ha realizado en el nivel simbólico. Para desarrollar el método para la descripción de la semántica-*t* se ha escogido como marco matemático la teoría de las categorías. La elección ha sido motivada porque permite usar un enfoque funcional de los procesos computacionales, y como se ha mencionado la semántica-*t* de un sistema es una función computacional. Al conjunto de métodos que se han desarrollado en la investigación para descripción de la semántica-*t* y que se

presentan en esta sección se los denomina *marco-S*.

### 8.4.1. La semántica- $t$ de los dominios de entrada

El problema a abordar es cómo describir en el DP la semántica- $t$  de un sistema computacional. Si el sistema computacional tiene una semántica- $t$ , estará ligado al sistema de entrada de señales. El sistema de entrada de señales tendrá un dominio de valores de los valores que son los datos que recibe el sistema. Así, el punto de partida es el dominio de datos de entrada que puede tener un sistema computacional, al que se denotará por  $\hat{D}_j$ . Por lo tanto, si un sistema computacional tiene un dominio de datos de entrada  $\hat{D}_j$ , se puede inferir la existencia de un entorno. Pero, también se puede inferir la existencia de otro elemento más. Ese elemento será una partición de los estados del entorno. La razón de que se infiera la existencia de un objeto que sea una partición es porque un dato de entrada no puede corresponder con un único estado del entorno desde el punto de la física fundamental, ya que eso implicaría que la cardinalidad del dominio fuera la misma que la del número de estados que tiene el entorno para que pudiera existir un isomorfismo, lo cual desde el punto de vista físico es irrealizable. Así, los elementos del dominio de entrada contendrán información solo de subestados del entorno, ya que cada subestado determina una clase de los estados del entorno. Téngase en cuenta que la partición del conjunto de los estados del entorno es definida a partir de los subestados, en base a que el subestado formará parte de todos los estados del entorno que contenga la clase. Formalmente eso implica que  $\hat{D}_j$  debe ser isomorfo a una partición del conjunto de estados del entorno. En base a eso se define un diagrama denominado, *diagrama estático de la semántica- $t$  de  $\hat{D}_j$* . Su definición es la siguiente:

**Definición 10** (Diagrama estático de la semántica- $t$  de  $\hat{D}_j$ ). *El diagrama estático de la semántica- $t$  de  $\hat{D}_j$  tiene tres objetos:  $\hat{D}_j$ ,  $E$  y  $E / \sim_{\tau_j}$ . El objeto  $E$  contiene todos los posibles estados del entorno del sistema computacional. El objeto  $S / \sim_{\tau_j}$  es una partición de  $E$  donde la relación de equivalencia viene dada por el transductor del sistema computacional, que es denotado por  $\tau_j$ . El diagrama tiene dos flechas:  $s_j$  y  $\tau_j$  y su representación gráfica es el diagrama 8.4.1.*

Pero la descripción que aporta el diagrama anterior es una descripción estática. Así, para lograr la descripción que permita comparar el comportamiento de la semántica- $t$  con la evolución del entorno, es necesario añadir un objeto denominado *dominio temporal* y denotado por  $T$ . El dominio temporal jugará un doble rol: 1) determina la evolución del entorno, y 2) permite comparar el comportamiento de la semántica- $t$  de diferentes sistemas. Por tanto, para describir el comportamiento de la semántica- $t$  de un sistema computacional se define el *diagrama dinámico de la semántica- $t$* . Ese diagrama se define de siguiente manera:

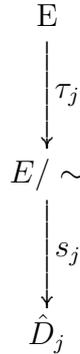


Diagrama 8.4.1: Diagrama estático de la semántica- $t$  de  $\hat{D}_j$ .

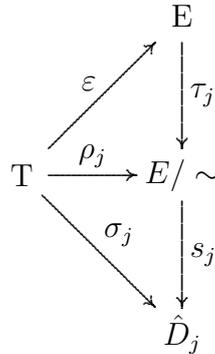


Diagrama 8.4.2: Diagrama dinámico de la semántica- $t$  de  $\hat{D}_j$

**Definición 11** (Diagrama dinámico de la semántica- $t$  de  $\hat{D}_j$ ). *Dada una  $\sigma_j : T \rightarrow \hat{D}_j$ , el diagrama del comportamiento de la semántica- $t$  tiene cuatro objetos :  $T$ ,  $\hat{D}_j$ ,  $E$  y  $E / \sim_{\tau_j}$ . El objeto  $T$  contiene todos los instantes del tiempo de manera ordenada. El objeto  $E$  contiene todos los posibles estados del entorno del sistema computacional. El objeto  $S / \sim_{\tau_j}$  es una partición de  $E$  donde la relación de equivalencia viene dada por el transductor del sistema computacional, que es denotado por  $\tau_j$ . El diagrama tiene cuatro flechas:  $\varepsilon$ ,  $s_j$ ,  $\tau_j$  and  $\rho_j$  y su representación gráfica es el diagrama 8.4.2.*

El siguiente paso es poder comparar descripciones de los sistemas. Pero antes de explicar la técnica que se ha diseñado se definirá qué son los *objetos fusionables*, ya que son la base de la técnica.

**Definición 12** (Objetos fusionables). *Los objetos fusionables de una lista de diagramas de la semántica- $t$  pertenecientes a una lista de sistemas son aquellos que son los mismos en todos los diagramas de la lista.*

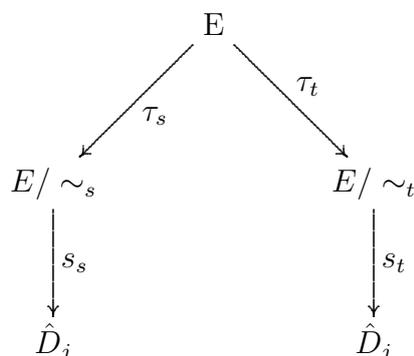


Diagrama 8.4.3: Diagrama dinámico de las semánticas- $t$  de un sistema experto con entrada de datos por consola y un sistema experto con sensores.

Entre los elementos fusionables de una lista de diagramas de la semántica- $t$  (del mismo tipo) de sistemas en el mismo entorno y ubicación temporal siempre serán fusionables los vértices  $T$  y  $E$ . Si los sistemas computacionales tienen los mismos transductores entonces  $E/\sim$  será un vértice fusionable también.

Considerando la definición anterior, la técnica para comparar las semánticas- $t$  de dos o más sistemas ligados a la que se denominará *fusión de diagramas de semánticas- $t$*  consta de dos pasos. El primero es localizar los objetos fusionables entre los diagramas. Una vez se tiene el conjunto de objetos fusionables, el segundo paso consiste en superponer los diagramas del comportamiento de la semántica- $t$  usando los objetos fusionables. El resultado de fusionar los diagramas del comportamiento de la semántica- $t$  puede ser un diagrama conmutativo o no conmutativo. Si el resultado es un diagrama conmutativo, las semánticas- $t$  son iguales. En caso contrario, las semánticas- $t$  son diferentes. Así, el diagrama resultante mostrará la semejanza de las semánticas- $t$  de los sistemas a los que pertenezcan los diagramas iniciales. La técnica se puede usar tanto con los diagramas estáticos como con los dinámicos de la semántica- $t$ , pero la lista de diagramas tiene que ser de un solo tipo.

La técnica de fusión de diagramas de la semántica- $t$  se puede usar para comparar el sistema experto que toma decisiones para el alto horno con la entrada de datos por teclado y el que su entrada de datos es mediante sensores de temperatura. Si se fusionan los diagramas estáticos de ambos, el resultado es el diagrama 8.4.3.

donde  $E/\sim_s$  es la partición creada por el sensor del sistema y  $E/\sim_t$  es la partición creada por el teclado.

El diagrama resultante es no conmutativo porque las particiones  $E/\sim_s$  y  $E/\sim_t$  son diferentes. El motivo es que por teclado se le puede dar un dato de entrada que no corresponda con la temperatura del entorno mientras que en el caso del sistema con sensores de temperatura, el dato de entrada siempre está ligado a la temperatura del

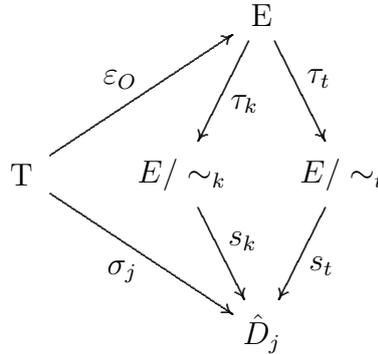


Diagrama 8.4.4: Diagrama dinámico de las semánticas- $t$  de dos sistemas expertos en el que el usuario introduce la temperatura que marca siempre el termómetro.

entorno. Por lo tanto, el diagrama no puede ser conmutativo. Así:

$$\exists e_x \quad s_t(\tau_t(e_x)) \neq s_s(\tau_t(e_x))$$

Esto muestra que los sistemas llevan a cabo distintos procesos computacionales a pesar de tener el mismo algoritmo.

Ahora bien, imagínese que se quiere comparar el proceso computacional de los dos sistemas cumpliéndose las condiciones que la persona siempre va a dar como dato de entrada el valor de temperatura que marque el termómetro del alto horno. Para hacer la comparación en este caso, se usarán los diagramas dinámicos de las semánticas- $t$  de los sistemas. La condición se puede expresar con la siguiente igualdad:

$$\forall t \quad s_t(\tau_t(\varepsilon_O(t))) = s_s(\tau_s(\varepsilon_O(t)))$$

Así, el diagrama resultante de la fusión de los diagramas dinámicos de la semántica- $t$  es el diagrama 8.4.4

### 8.4.2. La semántica- $t$ de las acciones de un sistema

En el apartado anterior se ha abordado la descripción de la semántica- $t$  de las señales de entrada de un sistema. Sin embargo, también es necesario describir como las acciones de un sistema que modifica el entorno adquiere una semántica- $t$ . Un sistema debe poder identificar cuál de sus estados son la causa de los cambios en su entorno; ese es un requisito fundamental para lograr sobrevivir en determinados entornos. Esos estados determinan las acciones que el sistema puede llevar a cabo, y es el conjunto de elementos sobre el que tiene que tomar sus decisiones. Además, este tema no puede ser ignorado porque se relaciona directamente con los valores de fasa reactiva relativos a la toma

de decisiones y el aprendizaje. Por lo tanto, este apartado, presenta la investigación realizada en este tema. Entre las acciones que realiza un sistema, la investigación se ha centrado son aquellas que actúan directamente sobre una extremidad o un motor. Así, se ha supuesto que los tipos de sistemas que se pretenden describir tienen un conjunto de estados que abren y cierran el flujo de energía en los circuitos del sistema para actuar sobre una característica observable del sistema. Por lo tanto, se asume la existencia de conjuntos de estados que activan los flujos de energía en los circuitos de un sistema exoactivo para controlar su exocomportamiento. A cada uno de esos conjuntos se le denominará *conjunto de acciones*, y se le denotará por  $A_j$ . Un conjunto de acciones será siempre un conjunto finito. Así,

$$A_j = \{a_1, \dots, a_{f(j)}\}$$

donde cada elemento  $a_i$  se denominará *acción*. La letra negra se usará para denotar el flujo de energía  $\mathbf{a}_i$  que es activado por el estado  $a_i$

La capacidad de un sistema para modificar el entorno con la notación usada en este capítulo se puede expresarse de la siguiente manera:

$$\varepsilon(t+1) = \eta(\varepsilon(t), \mathbf{i}(t))$$

donde  $\mathbf{i}$  representa la interacción generada por el sistema con el entorno en el momento  $t$ . La definición de  $\mathbf{i}$  es la siguiente:

$$\mathbf{i} : T \rightarrow \mathbf{A}_1 \times \dots \times \mathbf{A}_c$$

La función  $\mathbf{i}$  será abordada con más profundidad en el capítulo 9.

La definición de  $a_i$  es relativa al interés para estudiar sistemas en los cuales la capacidad del sistema para actuar en el entorno está relacionada con la capacidad para detectar el estado del entorno. Por supuesto, existe la posibilidad de que existan sistemas que puedan modificar características del entorno para el cual el sistema no tiene transductores que le permitan detectar el cambio que ha provocado, pero ese tipo de sistemas no serán seleccionados según la hipótesis de la libertad de movimiento vista en el capítulo 7. Por lo tanto, este último tipo de sistemas no ha sido considerado en la investigación.

Asumiendo todo lo que se ha mencionado, se ha realizado el siguiente razonamiento. Dado el conjunto de acciones  $A_j$  de un sistema exoactivo, su diagrama estático de la semántica- $t$  del dominio de entrada  $\hat{D}_j$ , entonces el hecho de que un cambio en  $E/\sim_j$  produzca un cambio en  $\hat{D}_j$  es decisivo para que el sistema tenga una semántica- $t$  de sus acciones. La razón de la importancia es porque implica que cada  $a_i$  pudiera ser reconocido por el sistema como:

$$a_i : \hat{D}_j \rightarrow \hat{D}_j$$

y al mismo tiempo  $a_i$  tiene asociado  $\mathbf{a}_i$ , que puede ser descrito por:

$$\mathbf{a}_i : E / \sim_j \rightarrow E / \sim_j$$

Por lo tanto, el resultado que se obtiene es que la interpretación de  $a_i$  sería la siguiente:

$$\mathcal{I}(a_i : \hat{D}_j \rightarrow \hat{D}_j) = \mathbf{a}_i : E / \sim_j \rightarrow E / \sim_j$$

Debe notarse que para que un sistema pueda tener una semántica- $t$  en sus acciones, debe cumplirse la siguiente condición. Dado que  $a_i$  es la causa del cambio, cuando el sistema establezca  $a_i$  en el instante  $t$  como su estado debe de producirse normalmente un cambio en  $t+1$  en el valor de  $\sigma(t+1)_j$  que debe ser igual a  $a_i(\sigma(t)_j)$ . Si esa condición no se cumple el sistema tendría un problema para establecer autónomamente su semántica- $t$ . Las causas de que no se cumpla pueden ser tres. Una puede ser porque  $a_i$  no produce instantáneamente la cantidad de energía  $\mathbf{a}_i$ . Otra puede ser porque  $\mathbf{a}_i$  no produce justo a continuación el cambio y hay un retraso. La tercera puede ser porque hay un retraso en la generación del valor en  $\hat{D}_j$  a partir del cambio en el estado del entorno. Si alguna de esas situaciones ocurriera, entonces el método para autodefinir  $a_i$  como  $a_i : \hat{D}_j \rightarrow \hat{D}_j$  se complica porque en la definición se tiene que incluir la definición de tiempo. Por ejemplo, si una motoneurona provoca un movimiento mediante la contracción de un músculo, pero el disparo de la neurona no causa instantáneamente el movimiento porque el potencial de acción viajara muy despacio por el axón hacia el musculo, la persona no podría en un primer momento conocer que fue ella la que lo provocó. Esto es porque la ventana de tiempo entre la activación de la orden y la información de que el movimiento ha sucedido permite que haya otros estados el sistema, por lo que cabría la opción de que uno de los siguientes estados fuera el desencadenante. Así, sería necesario un proceso de contrastación de estados para descartar estados.

### 8.4.3. El observador y el sistema de referencia

La idea de observador y sistema de referencia nació en la física. Un problema fundamental en física es estudiar cómo se relacionan las leyes físicas de dos observadores que están en marcos de referencia diferentes para determinar que, aunque ellos observen cosas distintas están observando el mismo fenómeno. El elemento clave para lograr demostrar que están observando el mismo fenómeno es el sistema de referencia que tiene asociado cada observador. Mira y Delgado incorporaron la idea de un sistema de referencia para un observador en sus propuestas para estructurar el paradigma computacional [391]. La idea es muy interesante pero desgraciadamente nunca llegaron a formalizarla ni desarrollarla en artículos. Puesto que en los dos apartados anteriores se han desarrollado múltiples conceptos para abordar la descripción de las semánticas- $t$  de los cálculos de los sistemas, en este se va a presentar una propuesta para formalizar los conceptos de observador y sistema de referencia.

En física se denomina observador a una entidad que puede hacer una medida. Por analogía, en el paradigma computacional se puede denominar observador<sup>3</sup> a un sistema que toma información y produce un comportamiento de acuerdo a la información que ha tomado.

Una simple descripción de la semántica- $e$  de un observador es:

$$D \times G \rightarrow A$$

donde  $D$  es un conjunto de descripciones del entorno,  $G$  son las metas que tiene el sistema, y  $A$  es el conjunto de acciones con las que el observador puede afectar al entorno. Así, las ranas son ejemplos de observadores, incluidas las del experimento de Sperry. Ahora bien, ¿cuál es el sistema de referencia del tipo de observadores que se ha definido? La respuesta puede extraerse de analizar dos observadores  $0_1$  y  $0_2$  donde  $0_1$  es una rana con la retina rotada y  $0_2$  es una rana que no ha sido operada; pero por lo demás, los dos observadores son iguales completamente. Supóngase que la semántica- $e$  de los dos observadores es la siguiente:

$$\begin{aligned} (d_1, g_1) &\mapsto a_1 \\ (d_1, g_2) &\mapsto a_2 \\ (d_2, g_2) &\mapsto a_1 \\ (d_3, g_1) &\mapsto a_3 \end{aligned}$$

En base a los experimentos de Sperry que se mencionaron en la sección 8.3, estos dos observadores realizan exocomportamientos diferentes, a pesar de tener la misma meta, las mismas acciones y el mismo conocimiento. Como se ha discutido en la sección 8.3, la razón de que realicen comportamientos diferentes es que tienen descripciones diferentes del entorno en el mismo momento de tiempo. Así, supóngase que la descripción de su cálculo mediante la semántica- $e$  son las siguientes. El sistema  $0_1$  tiene la descripción  $d_1$  y por lo tanto su cálculo es descrito por:

$$(d_1, g_1) \mapsto a_1$$

Sin embargo, el sistema  $0_2$  tiene la descripción  $d_3$  y su cálculo queda descrito por:

$$(d_3, g_1) \mapsto a_3$$

Este ejemplo muestra que los sistemas se comportan de acuerdo a las representaciones del entorno que ellos tienen, por lo tanto, la descripción que es determinada por la semántica- $t$  es un análogo a las coordenadas de objeto referentes a un sistema de referencia. Así, el *sistema de referencia* de un observador es su semántica- $t$ .

---

<sup>3</sup>Esta denominación difiere de la de Maturana y Varela, o Mira y Delgado, porque ellos además lo atribuyen la capacidad de inyectar conocimiento

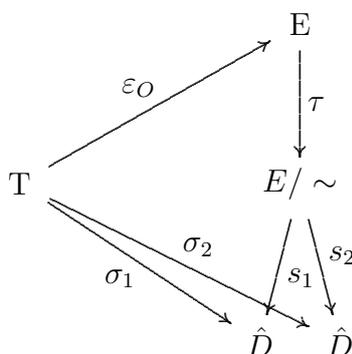


Diagrama 8.4.5: Diagrama dinámico de las semánticas- $t$  de los observadores  $0_1$  y  $0_2$ . El subíndice indica a que observador pertenece a función

Considerando todo lo anterior, surge un problema similar al formulado en la física sobre observadores en la TGE o las ciencias cognitivas. Dados dos observadores del entorno,  $0_1$  y  $0_2$ , surge el problema de describir la relación entre lo que cada sistema observa para producir sus exocomportamientos/comportamientos. A ese problema se le llamará, *el problema del acoplamiento de observadores*.

Para resolver el problema del acoplamiento de observadores se propone aplicar el marco- $S$ . Así, el problema del acoplamiento de observadores consistiría en hallar los diagramas de fusión de las semánticas- $t$  de los observadores.

En base a lo anterior, la solución del problema del acoplamiento de observadores para dos observadores  $0_1$  y  $0_2$ , siendo estos las dos ranas que se acaban de mencionar, es el diagrama 8.4.5.

El diagrama 8.4.5 es el resultado de la fusión de los diagramas dinámicos de la semántica- $t$  de los observadores, y muestra que ellos tienen diferentes sistemas de referencia. Por lo tanto, se puede decir, como en física, que observan situaciones diferentes a pesar de estar en el mismo entorno.

En este punto, surge la siguiente pregunta: ¿puede un sistema cuya semántica- $t$  ha sido modificada y no es compatible con su semántica- $e$  volver a lograr que su comportamiento sea adecuado al entorno? Esta pregunta está directamente relacionada con la discusión en la sección sobre por qué las ranas del experimento de Sperry no modificaban su comportamiento. Para analizar la pregunta, se crea el diagrama de fusión dinámico de las semánticas- $t$  y semánticas- $e$  de los dos sistemas 8.4.6. En el diagrama solo se fusionan  $T$  y  $E$  y se establecen flechas entre los vértices que podrían ser fusionados.

Para responder a la pregunta de si un sistema cuya semántica- $t$  ha sido modificada y no es compatible con su semántica- $e$  puede volver a lograr que su comportamiento sea adecuado al entorno, lo que se hace es estudiar el tipo de función que podría representar cada flecha entre los vértices que podrían ser fusionables. La interpretación

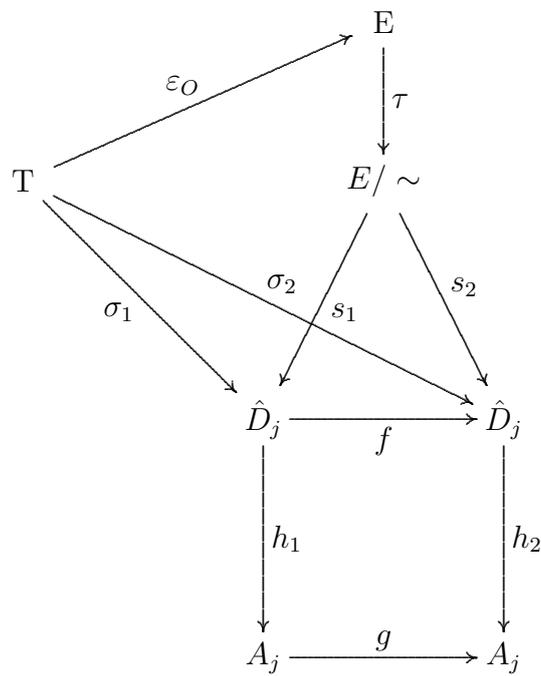


Diagrama 8.4.6: Diagrama de las semánticas- $t$  y semánticas- $e$  de dos observadores donde  $s_1$  es el proceso de detección de  $O_1$ ,  $h_1$  es la semántica- $e$  de  $O_1$ ,  $s_2$  es el proceso de detección de  $O_1$  y  $h_1$  es la semántica- $e$  de  $O_1$ .

de la pregunta que se ha planteado en el diagrama es la siguiente.

Si el observador  $O_1$  tiene una semántica- $t$  no modificada y el observador  $O_2$  tiene una semántica- $t$  entonces

$$s_1 \neq s_2$$

Si lo que se busca es que ambos observadores se comporten de la misma manera entonces se cumplirá que  $g$  es la función identidad de  $A_j$ . Es decir,

$$g = 1_{A_j}$$

En base a las condiciones que plantea el caso de la pregunta, se pueden deducir los dos siguientes teoremas sobre las opciones que permite el diagrama:

**Teorema 2.** *Dado el diagrama 8.4.6 si  $s_1 = s_2$  y  $g = 1_{A_j}$  entonces  $h_1 = h_2$ .*

**Teorema 3.** *Dado el diagrama 8.4.6 si  $s_1 \neq s_2$  y  $g = 1_{A_j}$  entonces  $h_1 \neq h_2$ .*

La razón para elegir esos teoremas entre todos los que pueden ser extraídos del diagrama 8.4.6 es porque muestran lo siguiente:

- Para describir el método por el que un sistema toma decisiones se necesita conocer la semántica- $t$  del sistema.
- Un sistema que tiene una semántica- $t$  diferente de la de otro necesita un método diferente para realizar el mismo comportamiento.

Hay otra cuestión destacable sobre el teorema 3, pero para comentarla primero se reformula el teorema interpretando los elementos del diagrama de la siguiente manera:

**Teorema 3 (reformulación).** *Dado un sistema que toma información del entorno para tomar decisiones, si para describirlo se usan solo estados del entorno y las decisiones realizadas por el sistema, hay más de una descripción de cómo el sistema toma sus decisiones.*

El teorema está relacionado con la lucha entre conductismo y ciencias cognitivas. Esa lucha la ganaron las ciencias cognitivas mediante experimentos [382]. Pero, el teorema apunta un nuevo elemento sobre esa lucha. El teorema dice que un investigador conductista puede encontrar una función que refleje cómo el sujeto toma decisiones en función de los estados del entorno; pero eso no implica que el sujeto esté haciéndolo como la función encontrada por el investigador conductista porque si las semánticas- $t$  de ambos son diferentes, el sujeto no estaría capturando el mismo estado del entorno que está capturando el investigador conductista.

En base a todo lo anterior, la conclusión final de investigar el problema del acoplamiento de observadores es que se si modifica la semántica- $t$  de un sistema, este tendrá que modificar su semántica- $e$  para recuperar su adecuación al entorno.

## 8.5. Análisis de los resultados

En la investigación presentada en este capítulo se han revisado las propuestas que se han realizado para estructurar el paradigma computacional. El paradigma computacional es el marco de las ciencias cognitivas, y en concreto de la IA y la psicología cognitiva, para plantear sus teorías. Pero, el paradigma computacional también es el marco del TGE, ya que la naturaleza de la fasa es computacional, y, por lo tanto, sus valores se describen computacionalmente. Lo que lleva a preguntarse si el paradigma computacional permite describir todos los valores de fasa posibles en la naturaleza. Las propuestas de Newell, Pylyshyn, Marr y Mira y Delgado para estructurar el paradigma computacional han estado enfocadas solo en las semánticas-*e* de los sistemas computacionales. Por ejemplo, Pylyshyn afirmaba lo siguiente:

“As I shall argue, unlike the symbol-process that characterize cognition cannot be described as a computation, that is by a purely symbol-manipulation function. Like all primitive operations of the functional architecture, the transducer fundamentally is a physical process;” [473] pag. 148:

En este capítulo se ha visto que, a pesar de la opinión de Pylyshyn, los transductores tienen una gran importancia para el paradigma computacional y que deben ser incorporados en la descripción computacional de los cálculos que llevan a cabo los sistemas para que la descripción sea completa. Existen dos razones de la gran diferencia de la estructuración que se propone en este capítulo del paradigma computacional con la de Pylyshyn. Una es que Pylyshyn considera que hay una diferencia entre el proceso físico y el proceso computacional. Sin embargo, en base a lo que se ha visto en el capítulo 5 esa distinción no se puede plantear en términos de entidades distintas, ya que el concepto de computación ha llegado a ser un concepto fundamental en física. Por lo tanto, no se puede afirmar que hay procesos físicos que no son computacionales. La otra razón es que Pylyshyn considera que cuando en la jerarquía de niveles se salta del nivel físico al nivel simbólico (algorítmico) se hace mediante un proceso de abstracción de las magnitudes físicas. Sin embargo, mientras ese proceso de abstracción conserva la semántica-*e* del sistema, no conserva la semántica-*t*. Esto es problemático, ya que saltar del nivel físico al nivel simbólico, aunque habrá una abstracción en algunas características de cómo los transductores funcionan, la conexión entre entorno y el proceso computacional debe de mantenerse en la descripción. Si eso no se hiciera ocurriría que sistemas que son diferentes, y que quedaría demostrado porque su comportamiento sería diferente ante entornos iguales, no podrían ser distinguidos. En la investigación también se ha visto que la propuesta más completa para estructurar el paradigma computacional entre las existentes es la de Mira y Delgado [393]. Sin embargo, se ha visto que tiene

limitaciones para determinados procesos computacionales en los que el sistema tenga semántica-*t*. Esto es debido a que la afirmación de que en el DP no existe semántica-*t* no se cumple en todos los sistemas computacionales. Esto ha conducido a investigar la realización de una propuesta para estructurar el paradigma computacional que permita describir todos los valores de fasa que existen en la naturaleza. Esa propuesta consiste en incorporar a la descripción computacional diagramas de la teoría de las categorías, ya que permiten usar un enfoque computacional funcional para mostrar explícitamente la semántica-*t* de los sistemas.

En base a los resultados obtenidos se ha propuesto la idea de que hay sistemas computacionales con diferentes semánticas. Esta idea implica la característica de que el DP cambia dependiendo del tipo de sistema computacional. Por lo tanto, una de las conclusiones centrales de este capítulo es que no se pueden obtener sistemas exoactivos con valores de fasa que generen exocomportamientos equivalentes al de los humanos solo desarrollando programas informáticos. La razón es que el tipo de *hardware* que posee el sistema computacional es importante para el proceso computacional. Esto no significa que el tipo de *hardware* debe ser un sistema nervioso bioquímico, al igual que un avión no vuela usando alas biológicas. Sin embargo, al igual que las alas de los aviones tienen que cumplir ciertos principios, el *hardware* de los sistemas tendrá que cumplir ciertos principios también. Por ejemplo, es necesario que el sistema computacional tenga transductores para que tenga una semántica-*t*. La idea de la necesidad de conectar el entorno con el sistema computacional ha sido defendida por Rodney Brooks[83]. Brooks ha propuesto que el robot tiene que tener un cuerpo que debe estar situado en el mundo real y experimentar el mundo directamente. La propuesta de Brooks se encuadraba en la propuesta de que el comportamiento inteligente puede ser generado sin una representación interna explícita con la que manipular [86], incluso sin un sistema de razonamiento explícito [85]. La investigación realizada en esta tesis muestra que la ubicación y la encarnación no son solo importantes para las arquitecturas de subsunción sino también son importantes para las arquitecturas simbólicas.

La investigación de la estructuración del paradigma computacional ha llevado a que este capítulo haya estado dedicado principalmente a la semántica. En palabras de Pylyshyn, “*the puzzle of meaning is the second hardest puzzle of cognition*” [473]. Los resultados presentados son un paso hacia una mejor comprensión del puzle, pero solo son eso, un paso. El puzle consta de múltiples preguntas complejas sobre lenguaje natural, y la investigación que se ha realizado solo ha abordado preguntas relacionadas a sensaciones y percepción. Además, hay otro tipo de semántica que no se ha abordado, la semántica intrínseca (aunque podría tratarse de diferentes tipos de semántica), a la que se denotará por semántica-*i*. Nada se sabe de este tipo de semántica; pero la prueba de su existencia es directa. La prueba está en cuando una persona sueña. A pesar de no tener señales de entrada en el cerebro desde el exterior, debido a que el cerebro se desconecta del exterior, el cerebro es capaz de generar un contenido que tiene el significado de un entorno real con objetos que tienen formas, colores y otras características. Además,

la semántica-*i* da contenido a sensaciones tales como sensaciones placenteras o dolorosas, que son diferentes de las semánticas-*t* porque son generadas por el propio cerebro, como prueba el síndrome de dolor central. A pesar de ser diferentes, la semántica-*t* y la semántica-*i* están muy relacionadas porque debe de haber una consistencia entre ellas cuando una persona está despierta. La semántica-*i* está directamente relacionada con el puzle más difícil que existe: la consciencia. Los resultados presentados en este capítulo no dan ningún paso para resolver ese puzle, y no hay ideas de cómo abordar este puzle desde el paradigma computacional. Hay pruebas experimentales que demuestran que la consciencia es causada por un proceso físico [316], pero se está muy lejos de una comprensión física completa de cómo se genera la consciencia. Un posible camino para entender la semántica-*i* es una estrategia evolutiva. La naturaleza solo selecciona las características que son una ventaja para los organismos. Por lo tanto, si la naturaleza mantiene la semántica-*i* en los organismos es porque aporta alguna ventaja. Así, un paso clave para entender como describir computacionalmente la semántica-*i* sería responder a la pregunta: ¿Cuál es la ventaja que conduce a la selección de organismos con semántica-*i* y no los que carecen de semántica-*i*?

La propuesta de usar el paradigma computacional para entender la mente de las ciencias cognitivas no ha sido universalmente aceptada. Como se mencionó al comienzo del capítulo, el ACA de Searle ha producido un acalorado debate sobre el paradigma computacional y las metas de la IA [471]. El argumento consiste fundamentalmente en imaginar un ser humano simulando ser una computadora. El ser humano está en una habitación en la que no puede ver nada del exterior, pero puede recibir y enviar documentos con símbolos. El ser humano dispone en la habitación de un libro con instrucciones que le dicen como rellenar otro documento a partir de los símbolos que encuentre en el documento que reciba. Tras recibir un documento, y una vez ha llevado a cabo todas las instrucciones que están en el libro, envía fuera de la habitación el documento que rellenó. Los símbolos que recibe en el documento son chinos y los que escribe en el documento de salida también. Así, un hablante de chino que ve las notas de entrada y de salida pensará que la persona que hay dentro de la habitación sabe chino, pero se sabe que la persona no entiende nada de chino. Así, Searle, concluye que, puesto que una computadora funciona de la misma forma que la persona lo hace en la habitación, manipulando símbolos mediante reglas, eso quiere decir que las computadoras no pueden entender genuinamente un idioma como lo hacen los seres humanos.

El ACA se ha considerado un argumento importante en contra del paradigma computacional para abordar la explicación de la mente humana. Esta controversia puede ser observada en los artículos “*Is the brain’s mind a computer program?*” de Searle [517] y “*Could a Machine Think?*” de Paul Churchland y Patricia Churchland [102]. Esos artículos son ampliamente conocidos porque defienden posturas opuestas sobre lo que la mente es y lo que la IA puede lograr. Sin embargo, aunque las preguntas de los títulos de cada uno de los artículos han sido consideradas equivalentes por

los autores de ambos artículos para presentar sus puntos de vista y argumentos; las dos preguntas no son equivalentes en base a la investigación que se ha presentado en este capítulo. Por esa razón, es interesante discutir ambas preguntas, pero desde los resultados que se han presentado aquí.

Primero, ¿es la mente un programa informático? Searle responde: no, no es un programa en base al AHC. Concretamente, Searle afirma lo siguiente: las mentes humanas tienen contenidos (semánticas) mientras los programas son formales (sintácticos), y la sintaxis por ella misma no es ni constitutiva, ni suficiente, para la semántica. La investigación que se ha presentado aquí concuerda con la afirmación de Searle. Así, él está en lo correcto porque un programa informático no puede contener una semántica-*t*. Por lo tanto, el AHC es correcto porque demuestra que un programa informático carece de una semántica-*t*. Cuando Searle imagina una persona dentro de la habitación, está considerando un sistema que tiene solo una semántica-*e* porque la situación no permite establecer una semántica-*t* para los símbolos chinos. Aquí, es importante notar que no es la misma situación cuando el ser humano está en la habitación china, que cuando una computadora comercial está en la sala china. La razón es que aunque ambos sistemas no tienen una semántica-*t*, Searle tiene una semántica-*i*, mientras que la computadora comercial no tiene. Esta diferencia es muy importante porque es el contraargumento del robot al AHC [154]. Cuando el programa es ejecutado en el robot no es el mismo caso que cuando Searle ejecuta el programa él mismo porque el robot no tiene una semántica-*i*. El argumento del robot propone que si en vez de una computadora con teclado, se tiene un robot con sensores el robot tendría semántica porque está conectado con el mundo. Searle contestó al contraargumento del robot que la persona que está en la habitación sigue obedeciendo reglas y no sabe lo que los símbolos significan. Searle afirma: “*he doesn't, for example, see what comes into the robot's eyes*” [519]p.196. La cuestión aquí es que mientras el problema cuando la persona están en la habitación es que no tiene una semántica-*t*, el problema del que se queja Searle en el contraargumento del robot es que no tiene una semántica-*i*. Searle no distingue entre los distintos tipos de semántica, pero la diferencia es muy importante porque el problema encarado cuando la persona está en la habitación y el problema del robot ejecutando el programa son dos problemas diferentes. Los recientes avances en retinas artificiales han permitido a personas recuperar parcialmente su vista. Así, si a la persona de la habitación le fueran sustituidas sus retinas biológicas por retinas artificiales, vería a través de los ojos del robot. Por lo tanto, el contraargumento del robot responde al problema original planteado en el ACR con la semántica-*t*, pero si lo que Searle quiere es una semántica-*i* (que es lo que él está argumentando cuando responde al contraargumento del robot) entonces el problema no es el paradigma computacional, sino la falta de conocimiento científico sobre la naturaleza de la semántica-*i* y la consciencia misma.

Searle ha usado el AHC para atacar la IA y las ciencias cognitivas, y los investigadores de IA y ciencias cognitivas han intentado demostrar que está equivocado. Sin embargo, si se admite que el AHC es correcto, al contrario de lo que Searle concluye,

no es necesario abandonar el paradigma computacional para entender mejor el proceso que lleva a cabo el cerebro. Desde los resultados presentados en este capítulo, se puede omitir la intención por la que Searle creó el AHC, y considerarlo una pista para lograr las metas de la IA y realizar descripciones computacionales más completas del proceso que lleva a cabo un cerebro.

La segunda pregunta es: ¿podría una máquina pensar? Los Churchland responden: sí, una máquina es capaz de pensar. Los Churchland argumentan: Primero, el axioma de Searle que afirma que la sintaxis por sí misma no es constitutiva ni suficiente para semántica podría estar equivocado porque no aporta pruebas suficientemente sólidas para defenderlo. En base a los resultados obtenidos en este capítulo, cómo se ha mencionado ya, Searle tiene razón, porque sintaxis y semántica son cosas diferentes. Por lo tanto, los resultados presentados aquí no apoyan el argumento de los Churchland. Sin embargo, despliegan un segundo argumento para su respuesta. Esos argumentan que una máquina podría pensar porque las computadoras digitales secuenciales son solo uno de los posibles tipos de computadora, y otro tipo de computadora podría evitar el AHC. Este segundo argumento para su respuesta sí es correcto en base a los resultados presentados. A pesar de que cuando los Churchland dan ese segundo argumento están pensando en computadoras paralelas y computación conexionista, la línea que se propone a partir de los resultados obtenidos es el de computadoras con semántica-*t* que evitarían el AHC.

Para ilustrar los resultados que apoyan que la investigación en Sistemas computacionales que tengan una semántica-*t* para evitar el AHC, se propone el siguiente experimento mental: Imaginemos un restaurante chino con una gran afluencia de personas y con gran capacidad. En este restaurante se cocinan todos los platos cada mañana. A continuación, se guardan en una cámara frigorífica, dónde se dejan los platos organizados en sus estanterías, de manera que en cada estantería se coloca un único tipo de plato. Todas las estanterías están organizadas según una secuencia en chino. En este restaurante tienen un problema, les falta una persona en el personal que tiene que estar trabajando en el servicio de comidas. Como no sería bueno para la imagen del restaurante tener a una persona que no fuera de origen chino en el restaurante sirviendo, han decidido contratar a un español que trabaje en la cocina, ya que allí nadie le ve y sólo tiene que coger el plato seleccionado, calentarlo y dejarlo en la ventanilla. Pero existe un problema, y es que el español contratado no sabe nada de chino. Así que han decidido dejarle un libro con reglas, mediante las cuales, cada orden que se le pasa, la traduce a los símbolos de la correspondiente estantería en la que se encuentra el plato. Aunque inicialmente el español contratado no sabe nada de chino, según pase el tiempo, va asociando la sintaxis de las notas que se le pasan con los platos que corresponden. Poco a poco comienza a ver símbolos que se repiten en todos los platos que contienen arroz. Así cuando ve ese símbolo mira a las estanterías que tienen platos con arroz. Según pasan los días, identificará más y más símbolos con los alimentos de los platos. Así, llega un momento en el que deja de necesitar utilizar el libro para localizar la es-

tantería correspondiente al plato para la orden. La razón por la que el español empieza a entender símbolos chinos es porque ha sido capaz de generar una semántica-*t* para una parte del chino.

Desde un punto de vista matemático, se ha afirmado que todos dominios de objetos matemáticos pueden ser codificados en  $\mathbb{N}$ . Así, una máquina de Turing tendría un poder computacional “total”, ya que todas las funciones efectivamente calculables entre dos dominios numerables cualquiera que fueran, podrían ser realizadas por una máquina de Turing. Sin embargo, esa afirmación no es correcta. La razón es porque cuando un dominio es codificado en  $\mathbb{N}$  y una máquina de Turing está ejecutando instrucciones para llevar a cabo la función, la máquina de Turing no está llevando a cabo la función original porque una función viene definida por su dominio y su codominio [187]. La función definida de  $\mathbb{N}$  a  $\mathbb{N}$  es una función diferente de la original. El hecho de que la máquina de Turing ejecute una función que codifica a la original solo prueba que la función original es computable. Uno podría pensar que esta es una discusión puramente matemática sobre la identidad de las funciones, pero no lo es. Hay funciones como las semánticas-*t* cuya identidad depende del dominio. Así, un programa ejecutado por una computadora comercial no puede tener un valor de fasa que pertenezca a la clase de los valores de fasa que tienen los seres humanos porque los ordenadores comerciales carecen de semántica-*t*. Aun así, podría proporcionar un modelo que ayude a estudiar cuestiones relacionadas con la semántica-*e* de los valores de fasa de los seres humanos. Desde los trabajos de Turing sobre computación, el punto de vista dominante ha sido considerar que un proceso computacional es independiente del substrato físico. Sin embargo, como han mostrado las investigaciones presentadas en este capítulo no son independientes del substrato. Una semántica-*t* solo puede ser llevada a cabo por un tipo específico de *hardware*.



## Capítulo 9

# TGE y Teoría Cognitiva de Condiciones de Verdad

En el capítulo 8 hemos visto que la semántica forma parte de los procesos computacionales en los que el sistema genera una respuesta en función de la información que capta del exterior. Si queremos desarrollar sistemas con comportamientos cognitivos complejos nos interesa investigar los valores de la fasa reactiva en los que la semántica juega un papel relevante.

Una propuesta para estudiar los aspectos semánticos de procesos computacionales es la teoría cognitiva de condiciones de verdad (TCCV)[371]. El origen de la TCCV fue desarrollar un marco matemático en el que definir matemáticamente el concepto de inteligencia para investigarlo formalmente. En el capítulo 2 se vio que el concepto de inteligencia es incapaz de proveer de una unificación de los comportamientos y de que ha recibido muchas críticas debido a las contradicciones que produce su diversidad de significados. El hecho de que la TCCV tuviera como intención inicial describir la inteligencia y que la inteligencia sea un concepto que se haya decidido descartar desde la TGE, no invalidan su uso para describir procesos de fasa reactiva.

La TCCV propone un marco matemático muy concreto y este puede ser usado para definir determinados valores de fasa. Concretamente, establece elementos para describir los procesos computacionales que toman información del entorno y determinan el comportamiento de un agente en el entorno. Por lo tanto, la TCCV encaja perfectamente como marco para describir los valores de la fasa que producen comportamientos complejos cognitivos.

La TCCV tiene tres funciones posibles. La primera es un marco donde desarrollar procesos simbólicos que reproduzcan habilidades cognitivas y en los que la semántica juega un papel importante. La segunda función es como marco en el que estudiar como la semántica está envuelta en procesos cognitivos a nivel simbólico. La tercera es como marco donde estudiar como estructuras sintácticas de arquitecturas cognitivas simbólicas pueden poseer semántica. La semántica resaltada por la TCCV es semántica-

$t$  vista en el capítulo 8. La TCCV considera que las condiciones de verdad no solo son una herramienta para dar significado a expresiones del lenguaje, sino que son elementos fundamentales en todas las características cognitivas que el cerebro humano posee.

Este capítulo realiza una introducción a la TCCV enfocada en los conceptos que subyacen a la teoría y limitando la notación y definiciones matemáticas. Primero se revisarán los principales elementos de la TCCV. Después se revisarán los mecanismos de tomas de decisiones que pueden ser desarrollados en la TCCV y se discutirán las diferencias de la TCCV con otros enfoques simbólicos para la toma de decisiones. Esta introducción tiene dos objetivos. Uno es proponer un marco matemático para investigar los valores de la fasa. El otro es que servirá para presentar el marco matemático en el que se han realizado investigaciones llevadas a cabo para la creación de robots multifuncionales basados en relaciones topológicas y que se presentarán en el siguiente capítulo.

## 9.1. La teoría cognitiva de condiciones de verdad

A principios del siglo XX, Hilbert propuso prescindir de la noción de verdad en su programa para fundamentar las matemáticas mediante un conjunto de axiomas finito y completo. Alfred Tarski reintrodujo la noción de verdad en matemáticas cuando dio una definición formal de verdad para formulas de lenguajes formales [336]. Entorno a la definición formal de la noción de verdad ha emergido una nueva rama de las matemáticas, llamada teoría de modelos, que estudia como los lenguajes formales describen clases de estructuras matemáticas. Esos dos tipos de elementos por un tipo de función llamada *interpretación*. Curiosamente, aunque Tarski estaba en contra de la noción de verdad en el lenguaje natural, su definición es la base para la propuesta de Donald Davidson para explicar el significado en el lenguaje natural. El propuso que una teoría finita del significado puede ser dada para un lenguaje natural mediante condiciones de verdad [116]. Las condiciones de verdad son formuladas mediante los esquemas T de Tarski explicados en el capítulo 8. Davidson argumentó que debido a que los seres humanos aprenden lenguaje empíricamente, el lenguaje natural debe ser estable en una forma finita, incluso si teóricamente es capaz de un infinito número de expresiones. Así, los significados de un número infinito de frases deberían estar basado en un número finito de axiomas. El concepto de semántica de teoría de modelos es importante también en psicología cognitiva (CP). El trabajo seminal de Philip Johnson-Laird dio lugar al nacimiento del campo denominado modelos mentales [248]. A pesar de la importancia de la semántica- $t$  en los campos mencionados, esa importancia no se ha reflejado en la investigación de las arquitecturas cognitivas. Ha habido poca investigación en el nivel simbólico de cómo la semántica- $t$  es relevante, o puede ayudar, para desarrollar agentes o robots autónomos. La TCCV es un marco matemático para hacer frente a ese vacío.

La principal idea detrás de la TCCV es crear algoritmos cuyo funcionamiento se ba-

se en que manejen símbolos que denotan elementos o características de los objetos del modelo. El primer elemento para lograr la meta anterior es una estructura matemática isomórfica a la realidad que los humanos perciben. El segundo elemento es un conjunto de lenguajes formales que debe permitir describir la estructura matemática. El significado de describir es que debe haber condiciones de verdad, la cuales determinen si una fórmula bien formada del lenguaje es verdadera o falsa en la estructura por medio de denotar elementos y propiedades de los elementos de la estructura.

Es importante considerar que la TCCV no propone que la estructura matemática conectada con los lenguajes formales por las condiciones de verdad corresponde con el universo físico. La selección natural no ha conformado cerebros porque refleje la estructura física del universo sino porque permite sobrevivir mejor y así tener más descendencia que otras, produciendo una reproducción diferencial.

Debido a que el formalismo matemático de la TCCV es demasiado extenso para ser explicado aquí, esta sección solo da una introducción con el uso de representaciones gráficas y evita largas definiciones matemáticas.

### 9.1.1. Estructuras de pasado presente futuro multivariadas con multiopción

Como se acaba de mencionar, el elemento de partida de la TCCV es una estructura matemática. Actualmente, la TCCV propone una clase de estructuras llamadas, estructuras de pasado presente futuro multivariadas con multiopción (PPFMM) [371]. Las estructuras PPFMM son estructuras de mundos posibles de Kripke, aunque son más complejas que las clásicas estructuras de mundos posibles usadas en lógicas intensionales. Una estructura PPFMM es una estructura anidada de mundos posibles porque cada mundo contiene a su vez una estructura de mundos posibles. Los mundos de una estructura PPFMM son denominados *perspectivas temporales*. La relación entre las perspectivas temporales es un orden total estricto. Una perspectiva temporal es a su vez una estructura de mundos posibles. Los mundos que contiene una perspectiva temporal son denominados *momentos de tiempo*. El número de perspectivas temporales es igual al número de momentos del tiempo. Esto hace que en cierta manera sea una estructura fractal. Cada momento de tiempo es un conjunto que contiene elementos que son tuplas. Cada una de esas tuplas es denominada *realidad*. Cada realidad es una terna, y dos posiciones de la terna determinan dos características: una es la *localización temporal* y la otra la *condición de realidad*. La localización temporal tiene tres posibles valores: pasado, presente y futuro. Todas las realidades de un momento del tiempo tienen como localizador temporal el mismo valor. La condición de realidad tiene dos posibles valores: existente e hipotética. Cada perspectiva temporal tiene solo un momento de tiempo con una realidad que sea existente y presente. La figura 9.1 muestra gráficamente los anteriores conceptos mencionados de una estructura PPFMM. Un importante punto

de la estructura PPFMM es que cuando se pasa de una perspectiva temporal a la siguiente, las características de los mundos cambian consistentemente con la flecha del tiempo psicológica. Por ejemplo, el momento del tiempo cuyas realidades están en el presente cambian al pasado cuando se pasa a la siguiente perspectiva temporal. En una perspectiva temporal, el momento siguiente al que sus realidades están en el presente y que está en el futuro cambia en la siguiente perspectiva temporal cambia la localización temporal de sus realidades a presente. Junto a la relación entre perspectivas temporales y la de momentos del tiempo hay una relación binaria entre realidades. Es importante notar que, aunque se usa la misma notación, los momentos del tiempo de diferentes perspectivas temporales son objetos diferentes. Se debería usar un índice denotando la perspectiva temporal a la que pertenecen. Sin embargo, la notación de realidades es general a todas las perspectivas temporales. Si en dos perspectivas temporales diferentes hay el mismo nombre para una realidad para cada perspectiva temporal ya que referencian la misma realidad.

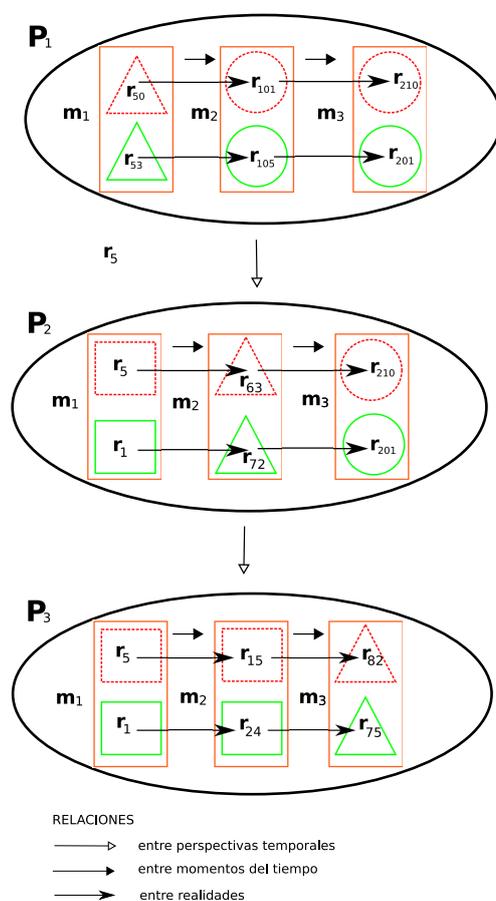


Figura 9.1: Representación gráfica de una estructura PPFMM. La estructura PPFMM mostrada aquí tiene tres perspectivas temporales:  $P_1, P_2, P_3$ . Cada perspectiva temporal tiene tres momentos temporales:  $m_1, m_2, m_3$ . Cada momento temporal tiene dos realidades. Las elipses representan perspectivas temporales. Los rectángulos son momentos temporales. Cuadrado = realidad pasado, triángulo = realidad actual, Círculo = realidad futuro. Verde/línea continua = realidad existente; rojo/línea discontinua = realidad hipotética

Antes se mencionó que una realidad es una terna, pero solo se han explicado dos posiciones de la terna. La otra posición de la terna contiene una estructura multivariada, la cual tiene tres tipos de dominios. Un tipo de dominios contiene identificadores para los objetos y sus partes. Específicamente, el dominio de objetos, denotado por  $\mathcal{V}$ , y un dominio que contiene las partes de los objetos,  $\mathbf{H}$ . El dominio que contiene los identificadores de las partes es denominado *conjunto esencia*. Los dominios son los mismos para cada realidad de una estructura PPFMM.

El segundo tipo de dominios es el de las propiedades que puede tener un objeto.

Los elementos de cada uno de esos dominios son los valores que cada una de las partes de un objeto puede poseer. Así, cada uno de esos dominios es llamado *propiedad*, y denotado por  $V_p$ . Cada  $V_p$  puede tener una estructura. Por ejemplo,  $V_p$  puede ser un espacio vectorial multidimensional. La existencia de dimensiones en  $V_p$  es importante porque, como se verá más adelante, además de considerar cambios en el valor de la propiedad es necesario considerar cambios en las dimensiones.

El último tipo de dominios contiene funciones que pueden ser ejecutadas por los propios objetos para alterar los valores de las propiedades y a las que se denominan *acciones*. Para cada propiedad y cada objeto, hay un dominio que contiene acciones. Cada objeto tiene asignado un conjunto de acciones.

Junto a los dominios, cada estructura multivariada contiene funciones y relaciones. Las funciones conectan los elementos de los diferentes dominios. Existe una función que asigna a cada objeto un conjunto de partes, y otra función por cada propiedad que asigna a cada parte un valor de la propiedad. Además, hay una función que asigna a cada objeto un conjunto de funciones. Cada objeto tiene siempre una acción asignada por cada propiedad y que no altera el valor de la propiedad. Esas acciones son denominadas como *acciones neutrales*. Si el conjunto de acciones asignadas a un objeto tiene asignadas otras acciones además de las acciones neutrales, el objeto es denominado *agente*.

Por su parte, las relaciones de la estructura multivariada determinan cuando un objeto influencia los valores de una propiedad específica de otro objeto. Esto significa que si un objeto,  $o_i$  está relacionado con otro objeto  $o_u$  en la propiedad  $p$ , entonces una acción del primero en la propiedad  $p$  afecta a los valores de la correspondiente propiedad del segundo objeto. Eso es, cuando una persona agarra una botella con su mano, la botella se mueve si la persona mueve la mano. Estas relaciones son denominadas *relaciones superconjunto momentaneo*.

Antes se mencionó que hay una relación binaria entre realidades. Esa relación esta conectada directamente con las acciones asociadas con cada objeto. La relación conecta una realidad de un momento del tiempo  $t$  con una realidad de otro momento del tiempo  $t'$  si los objetos de la realidad del momento  $t$  tienen acciones que cambian los valores que ellos tienen en  $t$  a los que ellos tienen en  $t'$ . Las acciones que conectan dos realidades son representadas con una tupla de acciones, llamada *interacción*. Una interacción contiene una acción para cada objeto y para cada característica. Las realidades pueden ser consideradas nodos de un grafo dirigido, y las aristas son determinadas por las interacciones. Las interacciones son siempre de la realidad de un momento del tiempo  $t$  a una realidad del momento de tiempo siguiente a  $t$ . Estos conceptos se muestran gráficamente en la figura 9.2.

Una perspectiva temporal puede ser vista como un grafo donde las realidades son nodos. Ese grafo cumple que dos nodos conectados pertenecen a dos momentos de tiempo distintos y consecutivos. A un camino del grafo se le denominará, *camino temporal*. Los caminos temporales son importantes porque son clave para los procesos de

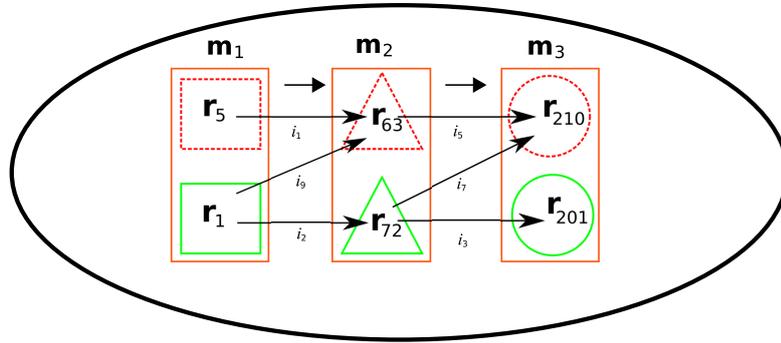


Figura 9.2: Representación gráfica de una perspectiva temporal. La perspectiva temporal muestra tres momentos temporales, cada uno de los cuales tiene dos realidades. Los rectángulos son momentos temporales. Cuadrado = realidad pasado, Triangle = realidad presente, Círculo = realidad futuro. Verde/línea continua = realidad existente, rojo/línea discontinua = realidad hipotética

aprendizaje y razonamiento en la TCCV.

La relación binaria entre realidades puede ser descrita como una función. Esa función, denominada *función productora* y denotada por  $\&$ , produce una realidad dadas una realidad y una interacción.

$$\begin{aligned} \& : R \times I &\longrightarrow R \\ (r, i) &\mapsto r \end{aligned}$$

Otra importante función que puede ser definida a partir de una estructura PPFMM es una función a la que se denominará *función interacción*, y se denota por  $\textcircled{i}$ . Esta función devuelve la interacción que ocurre en cada momento del tiempo. La función interacción es definida de la siguiente manera:

$$\begin{aligned} \textcircled{i} : T &\longrightarrow I \\ t &\mapsto i \end{aligned}$$

Usualmente, estaremos interesados en las acciones con las que un objeto concreto contribuye a las interacciones. Esa situación es particularmente relevante en las estructuras PPFMM donde solo hay un agente entre todos los objetos del entorno. En este caso, se añadirá un subíndice para indicar el agente,  $\textcircled{i}_u$ .

Usando las funciones  $\&$  y  $\textcircled{i}$ , se puede construir una función llamada *función sucesor*, y que se define de la siguiente manera:

$$\begin{aligned} Succ : (T \times R) \times I &\longrightarrow T \times R \\ \langle (t, r_x), i \rangle &\mapsto (t + 1, r_y) \end{aligned}$$

donde  $T$  es el conjunto de momentos del tiempo.

Se ha mencionado que las funciones  $\&$  y  $\textcircled{i}$  se obtienen de una estructura PPFMM; pero se puede ir en la dirección contraria, y definir una estructura PPFMM a partir de esas funciones.

El último elemento de las estructuras PPFMM al que se prestará atención en esta introducción es un dominio específico que contiene los estados de recompensa y castigo de un agente. Este dominio, denominado sistema de recompensa y castigo, es denotado por  $SRC$ . Los elementos del dominio de  $SRC$  son una tupla hecha de pares de la siguiente manera:

$$SRC = \{ \langle (s_1, c_1), \dots, (s_m, c_m) \rangle, \dots, \langle (c_1, c_1), \dots, (c_m, c_m) \rangle \} \cup \{ \emptyset \}$$

Cada par determina el nivel de recompensa para una sensación interna, como podrían ser el hambre, o la sed, en el caso del ser humano. El primer elemento del par determina el estado actual de la sensación de recompensa o castigo y el segundo es la máxima sensación de recompensa. Los elementos que componen el par están ordenados. Así que cuanto una mayor distancia del primer elemento del par respecto del segundo, mayor la aversión a la situación. Recompensa y castigo no existen en el mundo físico externo a los organismos. Esto podría llevar a considerar si estos deben añadirse en la estructura. Sin embargo, aunque no existen en el exterior, si existen en el cerebro a través de señales neuronales que los organismos integran en su representación de la realidad. Por lo tanto, sí forman parte del modelo que los organismos tienen de su entorno. Las recompensas primarias son fundamentales para dirigir el comportamiento en los animales. La estructura PPFMM captura ese hecho con la función  $\textcircled{\odot}$  que asigna un estado de recompensa y castigo a cada objeto. Esa función es definida de la siguiente manera:

$$\begin{aligned} \textcircled{\odot} : \mathcal{V} &\longrightarrow SRA \\ o &\mapsto \vec{s} \end{aligned}$$

Si  $\vec{s} \neq \emptyset$  el objeto es denominado sujeto.

Es interesante notar que si la función  $\textcircled{\odot}$  es definida en contradicción a las situaciones que permiten al sujeto sobrevivir, la función dirige al sujeto a su muerte. Por lo tanto, el propio proceso de selección natural produce la correcta definición de la función  $\textcircled{\odot}$  en la naturaleza.

Antes de terminar esta sección es interesante mencionar que la función interacción es el exocomportamiento de un sistema exoactivo  $u$ . Definir la función  $\textcircled{i}_u$  es equivalente a asignar una valor de fase al sistema exoactivo  $u$ . Un valor de fase aleatoria se formula mediante una variable aleatoria  $X$ . Esto se puede expresar de la siguiente manera:

$$\textcircled{i}_u(t) = C(X)$$

donde  $C$  genera valores cumpliendo la función de probabilidad que establece la variable aleatoria  $X$ .

En el caso de un valor de fasa recurrente se formula con una función iterada y su valor inicial. Así, se puede definir un valor de fasa recurrente mediante la siguiente expresión general:

$$\textcircled{i}_u(t) = C(f^t(x_0)) = \overbrace{f \circ \dots \circ f}^t(x_0)$$

Respecto a la fasa reactiva es necesario un aparato matemático más complicado para describir sus valores que los dos casos anteriores. Ese aparato matemático y la formulación de la fasa reactiva se verá en las siguientes secciones de este capítulo.

### 9.1.2. Jerarquía de lenguajes

La TCCV propone que un agente use una jerarquía de tres lenguajes formales para describir una estructura PPFMM. Cada uno de ellos es construido a partir del anterior en la jerarquía y cualifica más que el anterior. Este método para construir los lenguajes está basado en hechos de la neurociencia. En un sistema nervioso, un organismo registra toda la información a través de células receptoras, cada una está especializada para detectar una única clase de estímulo. Esas células registran el valor de un fenómeno físico en ese momento específico. Eso implica que cualquier otra información que el sistema nervioso contiene tiene que ser generada por el propio sistema nervioso. Así, la jerarquía de lenguajes es una réplica del proceso de generación de información a nivel simbólico. Esta característica es importante porque permite que la TCCV tenga un mecanismo para solucionar el problema del anclaje del símbolo en el nivel simbólico [209]. Un pequeño análisis sobre el problema del anclaje del símbolo es dado en la sección 10.6.

Los tres lenguajes formales de la jerarquía son denominados:  $L_{PPFMM}$ ,  $L_{PPFMM}^*$ , y  $LC_{PPFMM}$ . Todas las fórmulas de estos lenguajes referencian solo a una perspectiva temporal.  $L_{PPFMM}$  es la base de la jerarquía. Este lenguaje tiene una constante para cada elemento de cada tipo de la estructura PPFMM, pero no posee ni cuantificadores ni variables. Así,  $L_{PPFMM}$  describe con exactitud la estructura PPFMM pero no puede expresar ningún tipo de propiedad abstracta sobre la estructura ni sus elementos. El lenguaje  $L_{PPFMM}$  tiene cuatro tipos de fórmulas atómicas: Una es una completa descripción de un objeto, otra describe una relación de superconjunto momentáneo entre dos objetos, otra más expresa las acciones de las que un objeto es provisto, y la última describe el estado de recompensa y castigo de un objeto. El lenguaje, además de símbolos para denotar los valores de las propiedades, tiene símbolos para las propias propiedades. Cada fórmula atómica contiene un *indicador de realidad* y un operador denominado *situador temporal*. Hay dos indicadores de realidad:  $\varepsilon$ ,  $h$ . El indicador  $\varepsilon$  determina que la fórmula denota una realidad existente. El indicador  $h$  determina que la fórmula denota una realidad hipotética. Los situadores temporales son tres:  $E_{\downarrow}$ ,  $@_{\parallel}$ , y  $E_{\uparrow}$ .  $E_{\downarrow}$  expresa que la fórmula atómica denota una realidad de un momento de

tiempo del pasado desde la perspectiva temporal actual.  $@_{||}$  expresa que la fórmula atómica denota una realidad en el presente desde la perspectiva temporal actual.  $E_{|\downarrow}$  expresa que la fórmula atómica denota una realidad de un momento de tiempo en el futuro desde la perspectiva temporal actual. A partir de conectar fórmulas atómicas se pueden construir fórmulas complejas. Las fórmulas compuestas pueden ser construidas conectando fórmulas atómicas mediante conectores, que son conectores temporales. Hay cuatro conectores binarios:  $\blacktriangle$ ,  $\blacktriangleright$ ,  $\triangleright$ , y  $\triangle$ .  $\phi\blacktriangle\varphi$  determina que  $\phi$  y  $\varphi$  pertenecen a la misma realidad y momento del tiempo.  $\phi\blacktriangleright\varphi$  determina que  $\varphi$  pertenece a una realidad en el siguiente momento de tiempo en cual  $\phi$  está situado.  $\phi\triangleright\varphi$  determina que  $\varphi$  pertenece a un momento de tiempo sucesor de  $\phi$ , pero este no es el siguiente.  $\phi\triangle\varphi$  determina que  $\phi$  y  $\varphi$  describen un camino temporal y  $\varphi$  otro camino temporal diferente.

Entre todas las fórmulas bien formadas existen tres conjuntos de fórmulas importantes. El primer conjunto que se denomina *conjunto de momentos* contiene fórmulas cuyo único conector es  $\blacktriangle$ . El segundo es el *conjunto de series sencillas*, y cuyas fórmulas son aquellas del conjunto de momentos conectadas por  $\blacktriangleright$  o  $\triangleright$ . Cada fórmula del conjunto de series sencillas describe un camino temporal. El tercero es el *conjunto de series complejas*, y sus fórmulas son series sencillas conectadas por  $\triangle$ . El lenguaje  $L_{PPFMM}$  es completamente cuantitativo en el sentido de que para cada valor de una propiedad hay solo un símbolo en el alfabeto para denotarlo. En el nivel superior está  $L_{PPFMM}^*$ .  $L_{PPFMM}^*$  es una extensión de  $L_{MMPPF}$ . La cuestión importante sobre  $L_{MMPPF}^*$  es que contiene seis nuevos alfabetos. Estos alfabetos son denominados alfabetos de metainformación. Los elementos de dos de ellos,  $\sum_1$  y  $\sum_2$ , expresa diferentes estados cualitativos y de cambio en las propiedades de un objeto. El apéndice B presenta las definiciones de  $\sum_1$  y  $\sum_2$  y cómo se construyen las fórmulas atómicas correspondientes de  $L_{MMPPF}^*$  con esos alfabetos. Los elementos de los otros dos,  $\sum_3$  y  $\sum_4$  expresan diferentes estados cualitativos y cambios en una dimensión de una propiedad de un objeto. Los elementos de los dos últimos alfabetos de metainformación,  $\prod_1$  y  $\prod_2$ , expresan estados cualitativos y cambios las relaciones de superconjuntos momentáneos de un objeto con otro. La cualificación se realiza por medio de los valores de las propiedades y cómo estos valores cambian en relación al momento de tiempo anterior<sup>1</sup>. Los alfabetos de metainformación permiten la creación de nuevas fórmulas atómicas que describen las propiedades de los objetos cualitativamente.  $LC_{PPFMM}$  está en la cima de la jerarquía de lenguajes. Este lenguaje usa los alfabetos de metainformación de  $L_{MMPPF}^*$ , pero sus fórmulas atómicas también cualifican intervalos de tiempo. Esto es,  $LC_{PPFMM}$  se lleva a cabo una abstracción de la duración de los intervalos de tiempo. Hay cuatro tipos de fórmulas atómicas como en  $L_{PPFMM}$ . Las fórmulas atómicas pueden ser creadas con

<sup>1</sup>Es importante notar que la cualificación es relativa al momento de tiempo anterior y no al posterior. Aunque ambas opciones son posibles cuando observamos la estructura PPFMM, no es posible para el agente conocer la realidad futuro siguiente a la presente por razones obvias. Por ello, la cualificación se realiza usando la información de una realidad del momento de tiempo anterior.

elementos que pertenecen a conjuntos llamados categorías. Hay cinco categorías: objetos, patrones de objetos, condiciones de realidad, posicionadores temporales, y registros de estados. Estas categorías se definen formalmente de la siguiente manera:

- Objetos  $\mathcal{C}_O = \{o_1, \dots, o_z\}$
- Patrones de objetos  $\mathcal{C}_{PO} = \{\lambda o. \varphi_i, \dots\}$  donde  $\varphi_i$  es una fórmula que es verdad si el objeto cumple las características especificadas.
- Condiciones de realidad  $\mathcal{C}_{CR} = \{\varepsilon, h\}$
- Posicionadores temporales  $\mathcal{C}_{PT} = \{\downarrow |, ||, | \downarrow\}$
- Registros de estados  $\mathcal{C}_{RE}$  cuyos elementos son funciones  $\lambda$  que producen fórmulas atómicas. Hay tres tipos de funciones, que son las siguiente:
  - $\lambda tp \lambda o_i [tp | o_i | p | x] \quad x \in \sum_1 \cup \sum_2$
  - $\lambda tp \lambda o_i [tp | o_i | p | q | x] \quad x \in \sum_3 \cup \sum_4$
  - $\lambda tp \lambda o_i [tp | o_i | o_u | x] \quad x \in \prod_2$

donde  $p$  determina una propiedad,  $q$  determina una dimensión de una propiedad,  $tp$  es un posicionador temporal.

Cada elemento de una categoría se denomina *átomo*. Las fórmulas atómicas bien formadas de  $LC_{PPFMM}$  pueden ser generadas mediante la aplicación de  $\beta$ -reducción sobre registros de estados y los átomos para sus variables. Las fórmulas compuestas de  $LC_{PPFMM}$  tiene  $\blacktriangle$ ,  $\blacktriangleright$ ,  $\triangleright$ , y  $\Delta$  como conectores. Sin embargo, el significado de los conectores es diferente en  $LC_{PPFMM}$  del significado en  $L_{PPFMM}$  y  $L_{PPFMM}^*$  porque las fórmulas atómicas no son sobre específicos momentos de tiempo sino intervalos cualitativos de tiempo. Así, los conectores relacionan intervalos de tiempo.

### 9.1.3. Gramática reconocedora de condiciones de verdad

En el apartado anterior se ha descrito brevemente la jerarquía de lenguajes para describir una estructura PPFMM. La siguiente cuestión es cómo un agente, o robot, puede construir una fórmula para describir la estructura PPFMM. Un agente puede generar una fórmula bien formada de  $L_{PPFMM}$  que describe directamente la estructura usando la información registrada por sus sensores y concatenando y procesando las fórmulas atómicas constantemente. La fórmula de  $L_{PPFMM}$  que el agente genera usando la información de sus sensores es denominada *fórmula historial*. A partir de la fórmula historial, el agente puede generar de manera directa fórmulas de  $L_{PPFMM}^*$  para describir los estados cualitativos. Para crear una fórmulas atómicas de  $L_{PPFMM}^*$

del estado cualitativo de un objeto sólo es necesario una fórmula atómica de  $L_{PPFMM}$ , y dos fórmulas atómicas de  $L_{PPFMM}^*$  del mismo objeto de momentos de tiempo consecutivos para las fórmulas atómicas de  $L_{PPFMM}^*$  sobre cambios cualitativos del estado. Así, el agente solo necesita un proceso que sustituya cada valor de la propiedad por un elemento de los alfabetos de metainformación usando condiciones de verdad. Sin embargo, construir una fórmula de  $LC_{PPFMM}$  a partir de  $L_{PPFMM}^*$  es una tarea mucho más complicada. El motivo es que hay muchas fórmulas de  $L_{PPFMM}^*$  que están asociadas con una sola fórmula de  $LC_{PPFMM}$  debido a la . Esto se debe a la cualificación de los intervalos de tiempo que hace  $LC_{PPFMM}$ , y que conlleva que no hay una duración específica con una fórmula de  $LC_{PPFMM}$ . Así, se necesita un mecanismo que permita traducciones muchos a uno. Esto implica reconocer fórmulas de  $L_{PPFMM}^*$  y generar fórmulas de  $LC_{PPFMM}$ . Una gramática formal es un mecanismo que permite reconocer un número infinito de expresiones con un número finito de reglas. Sin embargo, esto no es suficiente porque se necesita que la capacidad de asociar expresiones de un lenguaje con expresiones de otro lenguaje. Afortunadamente, hay un mecanismo que ha sido estudiado y que contiene ambas características, los esquemas de traducción[12]. Los esquemas de traducción permiten traducciones de muchos a uno que es la característica que se necesita para resolver el problema. El esquema de traducción es usado para definir un mecanismo denominado gramática reconocedora de condiciones de verdad(GRCV). La GRV es un mecanismo que la TCCV usa para generar fórmulas de  $LC_{PPFMM}$  a partir de fórmulas de  $L_{PPFMM}^*$ . Pero antes de definir una TCCV, comencemos con las definiciones de traducción y de esquema de traducción.

Una traducción es un subconjunto de  $\Sigma^* \times \Upsilon^*$ , donde  $\Sigma$  y  $\Upsilon$  son alfabetos finitos. Un esquema de traducción es un sistema que generaliza la noción de una gramática libre de contexto para generar traducciones. Un esquema de traducción se define como  $G = (N, \Sigma, \Upsilon, R, S)$ , donde  $N$ ,  $\Sigma$ ,  $\Upsilon$  son secuencias finitas de *símbolos no terminales*, *símbolos de entrada* y *símbolos de salida*, respectivamente.  $N$  es disjunto  $\Sigma \cup \Upsilon$ .  $S \in N$ , es el símbolo de entrada.  $R$  es el conjunto finito de reglas. Una regla es un objeto  $A \rightarrow (\alpha, \gamma, \Pi)$ , donde  $A$  un símbolo no terminal, donde  $\alpha \in (N \cup \Sigma)^*$ ,  $\gamma \in (N \cup \Upsilon)^*$  y  $\Pi$  es una permutación.

Un esquema de traducción  $G = (V, \Sigma, \Upsilon, R, S)$  es sencillo si para todo  $A \rightarrow (\alpha, \gamma, \Pi) \in R$ ,  $\Pi$  es la permutación identidad. (i.e.,  $\Pi(i) = i$  para todo  $i$ .) Ya que la permutación de una regla es irrelevante para un esquema de traducción sencillo se puede eliminar de todas las reglas.

Una gramática reconocedora de condiciones de verdad(GRCV) es un esquema de traducción cuyas reglas de producción incorporan además un conjunto de asignaciones. Una GRV es un esquema de traducción sencillo en el lado derecho de las reglas de producción tiene a lo máximo un símbolo no terminal en  $\alpha$  y  $\gamma$ . Esto es porque las traducciones de  $L_{PPFMM}^*$  a  $LC_{PPFMM}$  puede ser hecha secuencialmente porque cada fórmula de  $L_{PPFMM}^*$  es una secuencia de fórmulas atómicas que describe un objeto en un camino temporal. Esta elección es debida a que el tiempo psicológico fluye en

una sola dirección. Pero una GRCV no es solo un esquema de traducción sencillo porque además cada regla contiene un conjunto de asignaciones. La razón es porque los elementos que son generados en la traducción son términos lambda. Así, junto con el término  $\gamma$  hay un conjunto de  $\Lambda$ . Dado lo anterior, la forma general de una regla de producción de una GRCV es la siguiente:

$$A \longrightarrow (\alpha, \gamma, \Lambda)$$

Así, siendo  $\gamma = \lambda x.M$  y  $\Lambda = \{x := c\}$  el siguiente proceso se lleva a cabo:

$$A \rightarrow \lambda x.M\{x := c\} \rightarrow_{\beta} M(c)$$

La primera transformación es generada por una regla de la GRCV, la segunda es mediante  $\beta$ -reducción.

El que la traducción genere términos lambda y asignaciones permite que aunque la traducción sea secuencial una decisión sobre la traducción pueda quedar a la espera de avanzar en la fórmula de entrada lo suficiente. Se debe notar que la forma en que un agente construye su fórmula historial está muy relacionada con cómo el agente puede determinar si una fórmula es verdadera o falsa. Dada una fórmula de  $LC_{PPFMM}$ , el agente debe verificar si la fórmula  $LC_{PPFMM}$  puede ser generada desde su fórmula historial para determinar si esta es verdad. Así, una fórmula bien formada de  $LC_{PPFMM}$ ,  $\varphi$ , es verdad en la estructura si una fórmula bien formada de  $L_{PPFMM}$ ,  $\phi$ , es verdad en la estructura y  $\varphi$  es la traducción de  $\phi$ .

## 9.2. Mecanismos para realizar decisiones de la TCCV

Hasta ahora hemos mostrado como la TCCV determina la manera en que un agente puede describir su entorno. Ahora, se abordará el tema de cómo la toma de decisiones puede ser lograda en la TCCV. La primera cuestión es que la TCCV establece una relación que conecta las estructuras PPFMM con la jerarquía de lenguajes. Una formulación sencilla es la siguiente

$$\textcircled{\mathbf{i}}_u(t) = r_t \pi^1(\vdash_u(\psi_u \varphi)) \quad (9.1)$$

donde  $\psi_u$  es la fórmula historial del agente  $u$ ,  $\varphi$  es la descripción de la estructura PPFMM,  $\vdash_u$  es una función que determina una secuencia de constantes que denota acciones del agente  $u$  para cambiar de  $\psi_u$  en el presente a  $\varphi$ ,  $\pi^1$  proyecta la primera constante de la secuencia producida por  $\vdash$ , y  $r_t$  es la realidad existente del momento del tiempo  $t$ . La ecuación expresa que la acción que el agente  $u$  lleva a cabo en cada momento del tiempo es la interpretación de las constante seleccionada por  $\pi^1$  de la secuencia calculada por  $\vdash_u$ .

La relación que se acaba de formular puede ser vista como una ecuación funcional si  $\vdash_u$  es considerada la incógnita. El problema es que no hay un método para determinar  $\vdash_u$ . Nótese que desde el punto de vista de la TGE  $\vdash_u$  es un valor de la fasa reactiva. La TCCV propone que una de las soluciones a la ecuación son las heurísticas de semántica cualitativa.

### 9.2.1. Heurísticas de semántica cualitativa

Las heurísticas de semántica cualitativa se basan en seis hechos. El primero de ellos es que los elementos de los alfabetos de metainformación son elementos definidos por una condición en un par de valores de una propiedad. Así, los alfabetos de metainformación podrían ser vistos como conjuntos de relaciones, y cada elemento de esos alfabetos de metainformación pueden ser vistos como una relación binaria entre dos valores de una propiedad.

$$R \subset V_p \times V_p$$

El segundo hecho es que las acciones pueden ser consideradas relaciones. Esto es debido a que una acción puede ser vista en una visión simplificada como

$$a : V_p \longrightarrow V_p$$

y cada función n-aria puede ser considerada como una relación n+1-ari. Así, una acción puede ser vista como:

$$a \subset V_p \times V_p$$

El tercer hecho es que dado que  $a$  provoca cambios siempre del mismo tipo y puede verse como una relación, entonces se cumplirá que hay una relación  $R$  para la cual:

$$a \subset R$$

El cuarto hecho es que una meta puede ser considerada un elemento de una relación binaria. Esto se debe a que una meta consiste en un cambio, el valor actual  $x$  por un valor  $y$ , y esos dos valores forman un par,  $(x, y)$ .

El quinto hecho es que los elementos de un alfabeto de metainformación como relaciones cumplen que son conjuntamente exhaustivas y disjuntas dos a dos. Así, hay una relación  $R$  y solo una para la cual:

$$(x, y) \in R$$

El sexto hecho es que la distancia que hay entre el estado actual y el objetivo de una meta es mayor, o igual, a la distancia que hay en los cambios que genera una acción.

Tomando en cuenta los hechos citados se puede concluir que :

$$(x, y) \in R \supset a$$

Así, la conclusión previa justifica que un agente puede lograr la meta  $(x, y)$  seleccionando la acción  $a$  etiquetada con  $R$ . Un lector podría afirmar que  $(x, y)$ ,  $R$  y  $a$  son elementos sintácticos. Sin embargo,  $(x, y)$  y  $a$  denotan elementos de una estructura PPFMM, y  $R$  es definida con elementos que denotan a su vez elementos de una estructura PPFMM. Esto implica que debe de estar en acuerdo con lo que denotan en la estructura. Si ellos no denotan elementos de la estructura, la conclusión será verdad solo si la forma de los símbolos es importante. Sin embargo, uno puede observar que es importante para sistemas biológicos o robots que haya un acuerdo entre lo que los símbolos denotan y como ellos se relacionan entre ellos. Si la definición sintáctica de una acción cambia un valor de temperatura por otro mayor pero cuando el agente usa esa acción ocurre lo opuesto, entonces hay un serio problema para que el sistema pueda comportarse adecuadamente. La relación entre los símbolos y lo que ellos denotan debe estar en acuerdo, si no la heurística no funcionará.

La idea de las heurísticas de semántica cualitativa está siendo investigada en el proyecto de investigación robots multifuncionales con nociones topológicas [379, 375] que forma parte también de esta tesis doctoral. Los resultados que trataremos en el capítulo 10 han mostrado que las heurísticas de semántica cualitativa permiten navegación a través de espacios desconocidos y abiertos[378].

El análisis de medios y fines(AMF) del General Problem Solver(GPS) [415, 143] trabaja de la misma manera las semánticas de heurística cualitativa. El AMF fue la primera heurística creada en IA. Los elementos de metainformación usados en las heurísticas de semántica cualitativa son equivalentes a las diferencias en el GPS. Sin embargo, Newell, Simon y sus colaboradores estuvieron siempre enfocados en diseñar un solucionador de problemas y solo dieron a las diferencias el papel de elementos clasificadores de operadores en las tablas de conexiones. Nunca consideraron que las diferencias podrían ser elementos de un lenguaje que permitiera a un agente describir su entorno, ni que semánticas de teoría de modelos estuvieran relacionadas con ellos. En realidad, se puede considerar AMF un elemento de la TCCV, pero las heurísticas cualitativas implican cuestiones de teorías de modelos que no son concebidas en el AMF.

### 9.2.2. Diferencias entre la TCCV y otros enfoques simbólicos

El enfoque lógico es uno de los más antiguos en IA. McCarthy lo propuso en el Symposium on Mechanization of Thought Processes celebrado en el National Physical Laboratory de Teddington en 1958. En IA, existen varios formalismos simbólicos como: cálculo de situación [70], cálculo de eventos [409], y el cálculo de fuentes [567]. Los dos

últimos resuelven el problema del marco. Por lo tanto, el lector podría estarse preguntando por la diferencia entre la TCCV y las otras propuestas. Para ver las diferencias es necesario tener en cuenta las diferencias entre los enfoques en lógica. En lógica matemática hay dos enfoques: teoría de la demostración y teoría de modelos. Teoría de la demostración es fundamentalmente sintáctico mientras que teoría de modelos es fundamentalmente semántico. Teoría de la demostración está enfocada en reglas y su validez es demostrada mediante una prueba de que no producen fórmulas que contradicen a los axiomas. Por otro lado, en teoría de modelos las reglas son válidas si ellas son semánticamente correctas en los modelos. Gödel demostró que el enfoque sintáctico no puede fundamentar las matemáticas, pero los sistemas deductivos de lógica de primer orden son consistentes y completos. En términos de computación, todo lo que es efectivamente calculable puede ser hecho con lógica de primer orden. Ese hecho es la razón por la que existen lenguajes lógicos de programación de propósito general.

El cálculo de situaciones, el cálculo de eventos y el cálculo de flujos son formalismos que usan el enfoque de teoría de la demostración. En estos formalismos se definen sistemas deductivos con un conjunto de axiomas, y se generan planes de acciones desde el enfoque sintáctico. Estos formalismos aplican resolución para lograr la meta. Resolución es una regla donde la semántica no juega ningún papel. Si se aplica iterativamente resolución permite decir si una fórmula es satisfacible y probar que una fórmula de lógica de primer orden es también insatisfacible. Así, estos formalismos solo consideran aspectos sintácticos. Sin embargo, la TCCV involucra el uso de semánticas porque como se ha mencionado los mecanismos para realizar decisiones están basados en las heurísticas de semántica cualitativa, y para confirmar la validez de sus reglas se las validada siempre con la estructura PPFMM. Esto conduce al hecho de que el cálculo de situaciones, el cálculo de eventos y el cálculo de flujos necesita mecanismos externos para solventar el problema del anclaje ya que no está contemplado en los propios formalismos por su enfoque sintáctico. Los predicados describen propiedades y estados de los objetos, los formalismos de los cálculos mencionados nunca dirigen como un robot o un agente pueden crear los predicados para describir su entorno a partir de la información de los sensores. Sin embargo, la TCCV contiene su propio mecanismo para manejar el problema del anclaje, la jerarquía de lenguajes. Por ejemplo, de la posición de un objeto, la jerarquía de lenguajes de la TCCV determina que un objeto se está moviendo o que está quieto.

## Capítulo 10

# Programa de investigación Robots multifuncionales mediante nociones topológicas

Se ha probado que los robots son herramientas útiles para los agentes de policía, cirujanos, y personal de limpieza. Sin embargo, en cada uno de esos casos los robots han sido diseñados para tareas específicas. Actualmente, se está estudiando la construcción de robots multifuncionales. Un robot multifuncional debe ser capaz de trabajar en entornos que requieren comportamientos flexibles como oficinas u hogares, y de interactuar directamente con los seres humanos en diversas actividades. Hasta ahora, muchos de los éxitos de la robótica están ligados al paradigma reactivo [80]. Sin embargo, parece difícil que el paradigma reactivo sea capaz de permitir el desarrollo de robots con la habilidad de actuar en entornos donde la tarea está cambiando constantemente. Desde el punto de vista de la TGE, el desarrollo de robots multifuncionales se plantea como la búsqueda de los valores de fase de robots que produzcan exocomportamientos equivalentes a los que realizan los seres humanos. La TGE afirma que el exocomportamiento de un organismo que provoca la construcción de un nicho en un entorno cuya órbita de estados no es cíclica ni abrupta respecto a la velocidad para responder del organismo, la fase del organismo será, al menos, de tipo reactiva. Pero además de eso, la TGE en su formulación actual no posee métodos para obtener un valor de fase específico. Así, aunque se puede plantear la investigación sobre robots multifuncionales desde la TGE el desarrollo de métodos solo puede ser llevado a cabo de manera constructiva. El primer indicio que se puede obtener para guiar la investigación es que el modo natural de comunicar tareas espaciales de los seres humanos es haciendo uso de relaciones espaciales. Así, un robot multifuncional recibiría continuamente tareas como:

*“Toma el paquete que está sobre la mesa y déjalo dentro del armario.”*

En base al hecho anterior, parece necesario que los robots multifuncionales que vayan a colaborar con seres humanos posean un alto grado de razonamiento espacial, y más específicamente, sobre relaciones espaciales cualitativas. La habilidad de un robot para moverse e interactuar en un entorno urbano depende de sus capacidades de percibir el entorno, autolocalizarse, navegar, planificar y procesamiento del lenguaje natural. Los procesos involucrados en esas tareas se comunican entre ellos a través de la representación espacial del robot. Los datos recogidos por los sensores (por ejemplo, [148, 447]) son usados para formar la representación del espacio, el método de autolocalización calcula la posición del robot en la representación, los métodos de navegación y planificación calculan los movimientos del robot a partir de la representación, y el método de procesamiento de lenguaje natural convierten las señales de voz en una sentencia sobre la representación.

Dependiendo de la complejidad del entorno y la tarea asignada a un robot, la arquitectura para la toma de decisiones debe dividirse en tres partes: ejecución, navegación<sup>1</sup> (o navegación local) y planificación (o navegación global). En cada uno de los niveles, los métodos pueden usar representaciones espaciales cuantitativas o cualitativas [528, 628]. La posible combinación de los niveles y los tipos de métodos ha producido diferentes tipos de arquitecturas, como muestra la figura 10.1. Las arquitecturas más investigadas en robótica son las de tipo I (mostrada en el diagrama). Este tipo de arquitectura es útil para aplicaciones de precisión industrial, pero no es apropiada para la interacción con humanos en la vida cotidiana. La causa de que no sea apropiada es que todas las representaciones espaciales de esa arquitectura son cuantitativas, por lo que es imposible para el robot comprender comandos como los que se dieron en el ejemplo de arriba. Esta limitación condujo a la investigación de métodos para la navegación cualitativa. [361, 299]. El desarrollo de métodos cualitativos ha producido la aparición de arquitecturas alternativas. Los primeros años de investigación en métodos cualitativos mostraron una gran dificultad para crear métodos de navegación cualitativa, lo que hizo que se formulara la “*the poverty conjecture*” [156]. Si esa conjetura fuera verdad, implicaría que la creación de una arquitectura de tipo III es imposible. Aunque, de acuerdo a algunos autores la investigación ha debilitado la conjetura [105]. Incluso así, hasta ahora, solo las arquitecturas de tipo II han sido desarrolladas exitosamente. Por lo tanto, el crear robots multifuncionales requiere de nuevos métodos de navegación cualitativa y un incremento de la capacidad de razonar sobre el espacio, lo que implica crear métodos nuevos que hagan uso de relaciones espaciales.

La formalización de relaciones espaciales ha sido investigada en otras disciplinas además de la robótica, como la lógica y la geografía. En geografía, uno de los más frecuentes modos para formalizar relaciones espaciales ha sido por medio de las relaciones

---

<sup>1</sup>En este texto, el termino navegación hace referencia solo a navegación local. Se usará el término planificar para referirse a la planificación global

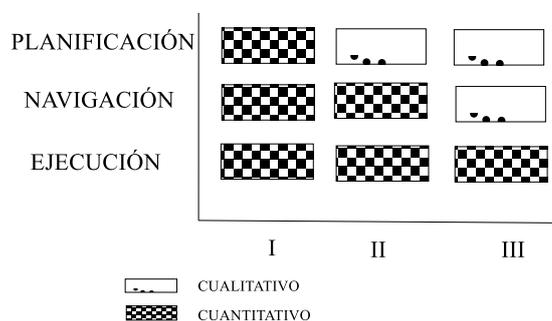


Figura 10.1: Un robot debe tomar decisiones a distintos niveles. Dependiendo del tipo de métodos usados por el robot en cada nivel se obtiene una arquitectura diferente para la toma de decisiones.

topológicas. Las relaciones topológicas se han usado en la creación de los sistemas de información geográfica (SSIG) para reconocer y analizar las relaciones espaciales que existen almacenadas en sistemas de información geográfica y, a su vez, permiten el análisis y definición de relaciones espaciales complejas. Además, la investigación psicológica indica que las relaciones topológicas son una parte importante del conocimiento espacial humano [484].

En el campo de la lógica y el paradigma simbólico de la IA se ha investigado en profundidad en los problemas de la satisfabilidad de configuraciones espaciales y la deducción de relaciones espaciales [477, 485, 484, 308, 511], la satisfabilidad de configuraciones de intervalos temporales y la deducción de relaciones temporales [15, 160]. En los comienzos del paradigma simbólico ya se investigaron algoritmos de planificación para cumplir configuraciones espaciales [153]. Sin embargo, solo recientemente se ha comenzado a investigar algoritmos para la toma de decisiones para navegación local [135].

Uno de los objetivos de la tesis ha sido desarrollar métodos que contribuyan a la creación de robots multifuncionales. Así, se inició un programa de investigación denominado robots multifuncionales mediante nociones topológicas (RMMNT). En el programa RMMNT se lleva a cabo una investigación para desarrollar métodos cualitativos que permitan que los robots puedan ser multifuncionales y puedan ser integrados en la vida cotidiana de los seres humanos. Esta investigación, se encuadra también como parte de la investigación de la TGE, ya que, desde el punto de vista de la TGE, la investigación que se lleva a cabo en el programa RMMNT consiste en la búsqueda de valores de fase que son capaces de producir exocomportamientos equivalentes a los que llevan los seres humanos para desenvolverse en sus entornos diarios. Este capítulo presenta las investigaciones realizadas en el programa RMMNT hasta el momento. La sección 10.1 presenta una introducción a la representación de espacios para métodos de navegación que sirve para encuadrar el programa RMMNT. La sección 10.2 presenta un resumen

de la heurística de semántica cualitativa topológica que ha sido el punto de partida del programa RMMNT. La sección 10.3 presenta la investigación que se ha llevado a cabo considerando dos posibles extensiones de las relaciones topológicas cualitativas. La sección 10.4 presenta una arquitectura para la navegación en espacios dinámicos y desconocidos abiertos, denominada arquitectura de navegación cualitativa topológica 1.0. En la sección 10.5 se presenta una heurística para la toma de decisiones sobre giros y como puede ser empleada para un sistema de coordenadas no cartesiano cualitativo. La sección 10.6 presenta la planificación de una segunda etapa del RMMNT centrada en la generación y procesamiento del lenguaje natural junto a los primeros resultados. La última sección realiza un breve análisis de los resultados obtenidos en el programa RMMNT.

## 10.1. Representación de espacios para la navegación

Ya se ha mencionado que las representaciones espaciales pueden ser divididos en dos clases: cuantitativas y cualitativas [628]. A su vez, los métodos para la toma de decisiones de la robótica pueden también ser divididos en cuantitativos y cualitativos, en función del tipo de representación espacial que usan. En esta sección, se van a examinar los métodos de representación usados en IA y su aplicación.

### 10.1.1. Métodos cuantitativos

En los métodos cuantitativos, el espacio se representa siempre como un espacio métrico. El problema cuantitativo de la navegación de robots móviles puede ser formulado de forma general como:

*Dado un punto de partida  $A$  alcanzar el(los) punto(s) de destino  $B$  ( $B_1, B_2, \dots$ ) utilizando su conocimiento [el del robot] y la información sensorial que recibe [el robot].*

Este problema ha sido ampliamente investigado en robótica. Los algoritmos desarrollados para resolverlo hacen que un robot se mueva desde unas coordenadas espaciales numéricas a otras [100]. La gran ventaja de los métodos de navegación cuantitativa es que con solo un nivel de representación espacial pueden ser llevadas a cabo diferentes tareas. Por ejemplo, la información capturada por los sensores puede ser usada directamente para construir una representación del espacio, el método de navegación local y global puede ser usado directamente para la representación del espacio, y el método de localización puede directamente calcular la representación de la posición del robot (por

ejemplo, mediante odometría o el sistema de posicionamiento global (GPS), navegación inercial, señales lumínicas, puntos de referencia). Dentro de la navegación cuantitativa, los métodos de representación espacial pueden ser divididos en dos clases: un robot un punto (URUP), y, un robot múltiples puntos (URMP).

El planteamiento de representación URUP se usa para navegar a un punto en el espacio con robots sin articulaciones y cuyas formas son constantes. En este planteamiento, el robot entero es representado como un único punto del espacio. En el caso de que un detallado mapa del entorno esté disponible, los algoritmos de función potencial [420], fuerzas [64], funciones armónicas [344] y métodos evolutivos [253, 252] entre otros son contemplados dentro de este planteamiento. En el caso de que no exista un mapa disponible para navegar, se suele usar algún algoritmo reactivo de tipo Bug [322, 255]. Los algoritmos de tipo Bug se basan en generar decisiones al reaccionar al entorno, lo que permite evitar obstáculos mientras se navega hacia un destino. La gran ventaja de este método de representación es que el algoritmo solo tiene que llevar a cabo cálculos para un punto dado del espacio. Obviamente, un robot no ocupa un punto del espacio. Evidentemente un robot móvil no ocupa un punto del espacio por lo que se suele, o ampliar los límites de los obstáculos para prevenir que haya choques, o cada punto del espacio equivale a la superficie del robot.

El planteamiento de representación URMP suele escogerse debido a que el robot con el que se trabaja tiene articulaciones y la complejidad del entorno hace necesario la planificación de trayectorias de las articulaciones [279, 98]. Así, cada parte móvil del robot es representada con uno o varios puntos del espacio [185, 37]. Si el robot debe ser representado con más de unos pocos puntos, calcular el movimiento de ese conjunto de puntos por el espacio tiene un gran coste computacional. Una de las propuestas que existe para tratar estos casos es el espacio de las configuraciones, donde cada punto del espacio representa una configuración. Así, lo que se hace es buscar una trayectoria del punto de la configuración de inicio al punto del espacio de la configuración final [321]. Pero construir el espacio de las configuraciones también tiene un coste computacional muy alto, y por ello, la mayoría de los algoritmos que usan el espacio de las configuraciones tienen una representación implícita que se explora para encontrar el camino [291].

### 10.1.2. Métodos Cualitativos

La diferencia fundamental entre los métodos cualitativos y los cuantitativos es el modo de representar el espacio. En los métodos cualitativos el espacio no es representado por un espacio métrico; pero cuál es en concreto el mejor modo en el que se debe representar el espacio no es una cuestión que esté resuelta. Incluso debido a que en los primeros años de investigación sobre métodos generales de razonamiento cualitativo espacial fueron infructuosos se enunció “the poverty conjecture” por Forbus, Nielsen y Faltings. Esa conjetura enuncia que:

“There is no purely qualitative, general-purpose representation of spatial properties.” [156]p. 418.

Aunque Cohn y Renz [105] consideran que la investigación realizada en los últimos tiempos parece refutar o al menos debilitar “the poverty conjeture”.

Entre los objetos de un espacio existen tres tipos de información espacial: topológica, de orientación y posicional. Las propuestas para construir representaciones del espacio cualitativas se basan en el uso de alguno de los anteriores tipos de información espacial. Algunas de las propuestas realizadas para construir una representación del espacio usando información topológica son  $RT_0$ [27, 201], el trazado topológico[480] y *Region Connection Calculus* (RCC)[477]; El trazado topológico está basado en el concepto de homeomorfismo [480]. Un mapa topológico agrupa posiciones que son equivalentes en base a las relaciones topológicas, lo que permite usar el mapa para realizar planificación de rutas[126]. El RCC es un cálculo lógico para razonar sobre configuraciones espaciales topológicas. Entre otras cuestiones, del RCC se ha estudiado su satisfacibilidad, su decidibilidad, la complejidad de distintas variantes, la composición de relaciones espaciales, y su mezcla con técnicas borrosas para manejar vaguedad [485, 483, 544, 307, 484, 306, 308, 511]. Pero, el RCC sólo permite descripciones espaciales estáticas, mientras que un robot se enfrenta a un espacio dinámico. En los últimos años se han realizado algunas propuestas para hacer descripciones dinámicas como *Spatio Temporal Constrain Calculus* [176], o para razonar sobre el movimiento como *Qualitative Trajectory Calculus* [609, 608, 610, 307, 121]. Pero estas propuestas son usadas para representar objetos en el espacio y el tiempo, y el problema que se ha investigado con estas propuestas es la satisfacibilidad de esas descripciones. Por lo que se desconoce el modo en que estas propuestas podrían ser útiles a la robótica. Usando información posicional se ha propuesto el panorama [505, 600], regiones orientadas[299], *Cardinal Direction Calculus* [310];y mediante información sobre distancia se ha propuesto las variables lingüísticas [266, 292] y los sistemas de distancias [104].

Prácticamente desde el nacimiento de la IA se ha investigado el problema de generar una secuencia de acciones para lograr una configuración espacial cualitativa entre objetos partiendo de otra configuración espacial cualitativa(problema del mundo de bloques, problema de las torres de Hanoi,...). El problema investigado puede enunciarse de forma general como:

*Dada la(s) relación(es) de partida  $S(S_1, S_2, \dots)$  entre los objetos  $[O, O']$  ( $[O_1, O'_1], [O_2, O'_2], \dots$ ) establecer la(s) relación(es) espacial(es)  $G(G_1, G_2, \dots)$  utilizando los operadores  $A_1, A_2, A_3, \dots$  y la base de conocimiento  $B$ .*

Este tipo de problema fue especialmente investigado desde los años 50 hasta los años 80. La descripción del mundo se hace con un lenguaje formal, en la mayoría de los casos

con lógica de primer orden. Algunas de las propuestas más importantes para resolver el tipo de problemas anterior son STRIPS[153], *cálculo de situación* [353] y *cálculo de eventos* [282]. *cálculo de situación* ha sido implementado en GOLOG [298] para el control de sistemas automáticos. Además, las variantes de GOLOG han sido creadas para controlar robots en la RoboCup [150]. Sin embargo, estos lenguajes están enfocados en fútbol y sus comandos solo manejan relaciones espaciales de un objeto dentro de una región [135]. Así, no existen hasta el momento métodos generales para traducir la información capturada por los sensores de los robots dentro de una representación de alto nivel con un grupo de relaciones espaciales completas.

Por otro lado, algunos expertos en el campo de la robótica consideran que la información topológica cualitativa es adecuada para la planificación; pero demasiado abstracta para que permita la realización de navegación [505]. Así, en robótica pueden encontrarse dos líneas de trabajo para el empleo de métodos cualitativos: jerarquías de lenguajes de descripción espacial y lenguajes menos abstractos que la información topológica.

- En las jerarquías de lenguajes espaciales se propone que el movimiento de un robot para desenvolverse en un entorno se debe dividir en distintos problemas, y cada problema se resuelve en un nivel de la jerarquía [497, 299, 361, 537, 280]. En estas jerarquías la información topológica cualitativa se usa para resolver problemas de planificación.
- Entre los métodos con lenguajes menos abstractos que la información topológica está panorama que usa información de orden [505]. Este método ha sido usado en la RoboCup [600, 599, 155].

### 10.1.3. Navegación basada en relaciones espaciales

Tras este resumen, se puede apreciar que hasta ahora las relaciones espaciales no han sido abordadas en la navegación de robots. Pero como se mostró en el ejemplo de tarea enunciada en lenguaje natural descrito al principio de este capítulo, las relaciones espaciales tienen una gran relevancia para lograr robots multifunción. Por lo tanto, en el camino para obtener robots multifunción es necesario investigar el uso de relaciones espaciales para la navegación. Respecto de las relaciones espaciales y la navegación, las dos primeras preguntas que cabe formular son:

- ¿Cómo se construye una descripción de relaciones espaciales a partir de una descripción cuantitativa?
- ¿Cómo se decide que acción realizar mediante una descripción de relaciones espaciales?

En base a esas dos preguntas se puede definir un nuevo tipo de problema de navegación, al que se va a denominar “el problema de navegación basada en relaciones espaciales”, y se enuncia de la siguiente manera:

*Dado un punto de partida  $A$ , establecer la(s) relación(es) espacial(es)  $G(G_1, G_2, \dots)$  entre los objetos  $[O, O']$  ( $[O_1, O'_1], [O_2, O'_2], \dots$ ) utilizando su conocimiento [el del robot] y la información sensorial que recibe [el robot]*

El problema de navegación basada en relaciones permite especificar lo que se quiere decir cuando se hace referencia a métodos de navegación cualitativa basada en relaciones espaciales y cuáles son los elementos que distinguen una arquitectura de tipo III.

#### **10.1.4. El programa robots multifuncionales mediante nociones topológicas**

El objetivo principal del programa RMMNT es desarrollar métodos que permitan la creación de robots multifuncionales que se incorporen a la vida cotidiana de los seres humanos, y propone que esos métodos estén basados en el uso de relaciones espaciales cualitativas. Su propuesta está motivada por el hecho de que los seres humanos usan relaciones espaciales cualitativas para comunicar tareas e información, ya que ese hecho hace que el uso de relaciones cuantificadas entre distancias no parecen suficientemente poderosas para permitir que un robot pueda participar adecuadamente en un entorno humano cotidiano. Así, el programa RMMNT está centrado en el desarrollo e investigación de métodos cualitativos para la robótica. El programa RMMNT también propone que la consecución de ese objetivo se haga en base a nociones topológicas. Desde el punto de vista de la TGE, el programa RMMNT consiste en la búsqueda de valores de fase equivalentes a los que poseen los seres humanos y que permiten desenvolverse en la vida cotidiana de los seres humanos. Obviamente, el objetivo final del programa RMMNT es un objetivo a largo plazo. Así, el programa consta de distintas fases en las que cada una tiene un objetivo que guía la investigación y los resultados que se buscan alcanzar. La primera fase del programa RMMNT tiene como objetivo la resolución del problema de navegación basada en relaciones mediante nociones topológicas. La segunda fase la generación y procesamiento de lenguaje natural con términos cualitativos. Las siguientes secciones presentarán los resultados obtenidos en estas dos fases, pero antes se discutirá el problema básico del que parte el programa RMMNT

#### **10.1.5. Representación mediante Espacios Topológicos**

En este apartado se van a presentar las definiciones formales del problema de navegación basada en relaciones espaciales y de su solución mediante nociones topológicas.

Al formalizarse mediante espacios topológicos se tienen dos niveles para describir el espacio. El primer nivel describe concretamente qué coordenadas del espacio están ocupando los objetos. Ese primer nivel es importante porque permite enlazar con los mecanismos de captación de un robot. Los sensores usan un lenguaje numérico, y, por lo tanto, a partir de la información de los sensores es posible la creación de una representación mediante la noción de espacio topológico métrico. El segundo nivel, que es más abstracto, es en el que se especifican las relaciones topológicas entre los objetos. Ese segundo nivel va a permitir que se formalicen las relaciones espaciales del problema que se ha enunciado.

En el problema la configuración inicial y la que perciba el robot por sus sensores serán descritas por el lenguaje del primer nivel, mientras que la configuración que se quiere lograr estará descrita por el segundo nivel. El robot ha de saber cuándo ha llegado a una configuración de primer nivel que cumple lo establecido por la que se quiere lograr. Así, es necesario establecer métodos de equivalencia entre ambos niveles. Por lo tanto, se van a presentar un conjunto de definiciones relativas a la formalización del problema en ambos niveles y la manera en la que se relacionan que parten de la noción de espacio topológico espacio topológico  $(X, \mathcal{T})$  donde  $X$  es un conjunto de elementos y  $\mathcal{T}$  una topología sobre  $X$  [262]. Las definiciones son las siguientes:

**Definición 13** (Estado Topológico Cuantitativo). *Dado un espacio topológico  $(X, \mathcal{T})$  y un conjunto  $O = \{o_1, \dots, o_m\}$ , al que se denominará conjunto de objetos, se llama estado topológico cuantitativo a  $e = \{(o_1, A_1), \dots, (o_m, A_m)\}$  donde  $o_i \in O$  y  $A_i \subset X$  con  $i \in \{1, \dots, m\}$ .*

Antes de definir la descripción de la configuración del espacio mediante el segundo nivel, es necesario hacer una definición de relación topológica entre objetos. Esta definición será el nexo entre los dos niveles de lenguaje topológico.

**Definición 14** (Relación Topológica entre Objetos). *Dado un espacio topológico  $(X, \mathcal{T})$  y un estado topológico cuantitativo  $e$  se dirá que la relación topológica entre dos objetos  $o_j$  y  $o_k$  es  $r_x$  en el estado  $e$  cuando se cumpla que  $A_j r_x A_k$*

Mediante el concepto de la relación topológica entre objetos se define estado topológico cualitativo de la siguiente manera:

**Definición 15** (Estado Topológico Cualitativo). *Dado un espacio topológico  $(X, \mathcal{T})$  y un conjunto  $O = \{o_1, \dots, o_m\}$ , al que se denominará conjunto de objetos, se llama estado topológico cualitativo al conjunto  $\{(o_1 r_x o_2), \dots, (o_1 r_v o_m), (o_2 r_{x'} o_1), (o_2 r_u o_3), \dots, (o_m r_v o_{m-1})\}$*

Hay que darse cuenta que un objeto  $o_j$  siempre tiene la relación topológica de igual consigo mismo en cada estado, y por lo tanto está información no hace falta hacerla explícita en el estado topológico cualitativo.

Una vez se han definido estado topológico cuantitativo y cualitativo se puede formalizar un problema de navegación basada en relaciones espaciales mediante nociones topológicas de la siguiente manera:

**Definición 16** (Problema de Navegación Topológica Cualitativa ). *Dado un espacio topológico  $(X, \mathcal{T})$ , un conjunto de objetos  $O$  y un conjunto de funciones  $F = \{f_1, \dots, f_s\}$  donde  $f_i : X \rightarrow X \quad i \in \{1, \dots, s\}$  , se denominará problema de navegación topológica cualitativa a la tupla  $((X, \mathcal{T}), O, e_1, F, R, \Omega, \Lambda)$  donde:*

- $e_1$  es un estado topológico cuantitativo, al que se denominará estado inicial.
- $R$  es un conjunto de relaciones topológicas  $R = \{r_1, \dots, r_n\}$  .
- $\Omega = \{o_j r_x o_k, \dots\}$  es el conjunto de relaciones topológicas entre objetos que se tienen que cumplir.
- $\Lambda$  es una función que asigna a cada objeto un conjunto de funciones que representan las acciones que es capaz de realizar ese objeto.

$$\Lambda : O \rightarrow \mathcal{P}(F)$$

*El conjunto de funciones que asocia la función  $\Lambda$  a cada objeto es siempre finito.*

De entre todos los problemas posibles, la investigación se centró inicialmente en los que tienen como espacio topológico a  $\mathbb{Z}$  con la topología del orden. El interés por  $\mathbb{Z}$  es debido a que es un espacio no denso. Los mecanismos de captación de los robots recogen información discreta sobre su entorno (un número finito de sensores en el sonar o láser, un número finito de píxeles en la cámara,...). Por otro lado, el que  $\mathbb{Z}$  no tenga ni ínfimo, ni supremo, permite que pueda tomarse como una generalización de todos los casos de espacios finitos no densos. Además, analizar el problema de navegar por un espacio de una sola dimensión tiene la ventaja de tener menos complejidad que analizarla sobre un espacio de varias dimensiones, y por lo tanto permite hacer un mejor análisis del problema.

Una vez se ha definido una representación formal de un problema de navegación a partir de la noción de espacio topológico, la siguiente cuestión que se ha abordado es definir en qué consiste una solución y su representación formal.

**Definición 17** (Solución Topológica Cualitativa). *Dado un problema de navegación topológica cualitativa  $((X, \mathcal{T}), O, e_1, F, R, \Omega, \Lambda)$ , se denominará solución topológica cualitativa a una secuencia  $\langle ((o_1, f_{\mu_1(1)}), \dots, (o_m, f_{\mu_m(1)})), \dots, ((o_1, f_{\mu_1(z)}), \dots, (o_m, f_{\mu_m(z)})) \rangle$  que cumple:*

- $\forall (o_j r_x o_k) \in \Theta \quad f_{\mu_j(z)}(\dots, (f_{\mu_j(1)}(A_j))) \quad r_x \quad f_{\mu_k(z)}(\dots, (f_{\mu_k(1)}(A_k)))$   
siendo  $(o_j, A_j), (o_k, A_k) \in e_1$ .

La función  $\mu_k(x)$  es la función que determina cual de todas las posibles funciones que tiene asignado el objeto  $o_k$  por la función  $\Lambda$  se usa en la posición  $x$  de la secuencia. Se entiende que las funciones se han ordenado bajo algún criterio que permite indexarlas puesto que son un número finito.

Mediante las definiciones formales anteriores ahora es posible enunciar el objetivo principal de la investigación, que es encontrar un método de razonamiento automático que permita encontrar la solución topológica a un problema de navegación topológica cualitativa. Es decir, si se denota por  $\vec{p}$  un problema de navegación topológica cualitativa, se busca una heurística  $h$  tal que genere una solución topológica cualitativa  $\vec{s}$ , de manera que  $h(\vec{p}) = \vec{s}$ .

## 10.2. La heurística de semántica cualitativa topológica

Para que un agente pueda establecer una relación espacial, no se pueden usar métodos de razonamiento debido a que son métodos que permiten determinar el tipo de cambios que se deben llevar a cabo para establecer un objetivo; pero no permiten determinar las acciones que hay que llevar a cabo. Es decir, los métodos de razonamiento son útiles para planificar, pero no para navegar. Así, para navegar es necesario usar métodos de toma de decisiones. En el caso del programa RMMNT, el método de toma de decisiones debe determinar las acciones que se tienen que llevar a cabo para establecer una relación espacial. En base a lo anterior, la investigación del programa RMMNT ha partido de unos resultados obtenidos sobre el desarrollo de una heurística de semántica cualitativa para la toma de decisiones mediante nociones topológicas [372]. Los resultados obtenidos permitían tomar decisiones para un espacio no denso unidimensional con la topología del orden, como por ejemplo  $\mathbb{Z}$ . Como se ha mencionado, la razón del interés en ese tipo de espacios es porque los sistemas que capturan información en los robots están físicamente limitados en precisión y tienen un número finito de sensores.

El primer paso del programa RMMNT fue revisar en profundidad los resultados que se habían obtenido para establecer relaciones. En la revisión se descubrió que, aunque en principio solo se habían considerado tres posibles casos determinados por los tamaños del objeto y el agente para tomar decisiones, en realidad existen seis casos debido a que el espacio que se usa es discreto. Puesto que los resultados iniciales estaban basados en la búsqueda de una heurística de semántica cualitativa para espacios topológicos de las que se ha hablado en el capítulo 9, se decidió denominar al método de toma de decisiones *heurística de semántica cualitativa topológica* (HSCT). A continuación, se presentará la HSCT.

El elemento fundamental de la HSCT es el grafo lineal de razonamiento topológico

(GLRT). Un GLRT es una tupla cuyos elementos pertenecen al conjunto de las *relaciones topológicas relativas* (RRTTRR). Las RRTTRR son un conjunto de relaciones topológicas diferentes de las que se usan en los modelos “9-intersection model” o RCC-8. Cuando en esta tesis se hable de las relaciones topológicas del “9-intersection model” o RCC-8 se las denominará como las relaciones topológicas clásicas. Las relaciones topológicas clásicas son 8. Concretamente, se tratan de *disjuntos*, *pegados*, *solapados*, *cubierto*, *cubre*, *contenido*, *iguales* y *contiene*. Se suele denotar el conjunto de relaciones topológicas clásicas por  $R_8$ . El conjunto de RRTTRR, al que denotara por  $S_{13}$ , tiene trece elementos. Así,  $S_{13} = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}, s_{11}, s_{12}, s_{13}\}$  donde a  $s_1$  se le denomina *disjuntos-0*, a  $s_2$  se le denomina *pegados-0*, a  $s_3$  se le denomina *solapados-0*, a  $s_4$  se le denomina *cubierto-0*, a  $s_5$  se le denomina *cubre-0*, a  $s_6$  se le denomina *contenido*, a  $s_7$  se le denomina *iguales*, a  $s_8$  se le denomina *contiene*, a  $s_9$  se le denomina *cubierto-1*, a  $s_{10}$  se le denomina *cubre-1*, a  $s_{11}$  se le denomina *solapados-1*, a  $s_{12}$  se le denomina *pegados-1*, a  $s_{13}$  se le denomina *disjuntos-1*. Las relaciones del conjunto de las relaciones topológicas relativas son exhaustivas conjuntamente y disjuntas dos a dos.

El uso de las RRTTRR frente a las relaciones topológicas clásicas proviene de que se vio que la capacidad descripción de las relaciones topológicas clásicas era insuficiente para la toma de decisiones para la navegación [372, 373, 379]. El incremento en el número de relaciones aporta la capacidad para permitir romper ambigüedades en la descripción de estados del espacio.

Como se ha mencionado, en la revisión de los resultados que se tomaron como punto de partida, se vio que en vez de tres casos para asignar un GLRT existen seis casos. El conjunto de reglas para asignar el correspondiente GLRT a un agente y un objeto es relativo a sus tamaños y formas. Para un espacio no denso con una topología del orden están en la tabla 10.4.

Caso	Subcaso	id	Grafo lineal de razonamiento topológico
$ A  >  B $	$ A  > ( B  + 1)$	1	$\langle s_1, s_2, s_3, s_5, s_8, s_{10}, s_{11}, s_{12}, s_{13} \rangle$
	$ A  = ( B  + 1)$	2	$\langle s_1, s_2, s_3, s_5, s_{10}, s_{11}, s_{12}, s_{13} \rangle$
$ A  <  B $	$ A  < ( B  + 1)$	3	$\langle s_1, s_2, s_3, s_4, s_6, s_9, s_{11}, s_{12}, s_{13} \rangle$
	$( A  + 1) =  B $	4	$\langle s_1, s_2, s_3, s_4, s_9, s_{11}, s_{12}, s_{13} \rangle$
$ A  =  B $	$ A  =  B  > 1$	5	$\langle s_1, s_2, s_3, s_7, s_{11}, s_{12}, s_{13} \rangle$
	$ A  =  B  = 1$	6	$\langle s_1, s_7, s_{13} \rangle$

Tabla 10.1: Condiciones entre dos objetos para asignar un GRLT para un espacio topológico no denso y unidimensional.  $|X|$  es el tamaño de  $X$ .

Las RRTTRR son calculadas en  $\mathbb{Z}$  mediante el formalismo *matriz de proposiciones*

de orden, y denotada por  $P^{\cong}(X, Y)$ , que consiste en la siguiente matriz :

$$P^{\cong}(X, Y) = \begin{pmatrix} \text{mín}(X) < \text{mín}(Y) & \text{mín}(X) < \text{máx}(Y) \\ \text{máx}(X) < \text{mín}(Y) & \text{máx}(X) < \text{máx}(Y) \\ \text{mín}(X) = \text{mín}(Y) & \text{mín}(X) - 1 = \text{máx}(Y) \\ \text{máx}(X) + 1 = \text{mín}(Y) & \text{máx}(X) = \text{máx}(Y) \end{pmatrix}$$

donde  $X$  es el espacio ocupado por el agente e  $Y$  el espacio ocupado por el objetivo.

Cada uno de los elementos de la matriz de proposiciones de orden es una proposición que toma un valor del conjunto  $\mathbb{B} = \{0, 1\}$ , dependiendo de si la proposición es falsa (0) o verdadera (1). En la tabla 10.2 se muestra la definición de cada RTR mediante la matriz de proposiciones de orden.

La HSCT requiere para su implementación las estructuras de datos siguientes:

- Una tabla de medios y fines con las acciones del agente etiquetadas por una de las tres relaciones siguientes:  $\{<, =, >\}$
- Un conjunto de reglas para asignar el correspondiente GLRT para cada agente y objeto que se puedan definir en el espacio.

Usando las anteriores estructuras de datos, la HSCT se implementa mediante los siguientes pasos:

1. Se elige un GLRT en base al tamaño de los objetos  $o_A$  y  $o_R$ .
2. Se calcula la RTR,  $s_c$ , entre los objetos  $o_A$  y  $o_R$  a partir de sus posiciones actuales y se almacena el número  $n_{s_c}$  asociado a  $s_c$  en el GLRT.
3. Se obtiene el número,  $n_{s_t}$ , asociado con el nodo,  $s_t$ , de la RTR que es el objetivo a establecer entre los objetos  $o_A$  y  $o_R$ .
4. Se almacena la relación de orden que se cumple entre  $n_{s_c}$  y  $n_{s_t}$  de las siguientes:
  - $n_{s_c} < n_{s_t}$
  - $n_{s_c} = n_{s_t}$
  - $n_{s_c} > n_{s_t}$
5. Se consulta la tabla de medios y fines usando la relación de orden que se cumplía entre  $n_{s_c}$  y  $n_{s_t}$  para recuperar la acción que esté etiquetada con la relación de orden mencionada. Esa acción es la decisión que debe llevarse a cabo.

Disjuntos-0	Pegados-0	Solapados-0
$P^< = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$	$P^< = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}$	$P^< = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$
Cubierto-0	Cubre-0	Contenido
$P^< = \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}$	$P^< = \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$	$P^< = \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$
Iguales	Contiene	Cubre-1
$P^< = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$	$P^< = \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$	$P^< = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}$
Cubierto-1	Solapado-1	Pegados-1
$P^< = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$	$P^< = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$	$P^< = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$
Disjuntos-1		
$P^< = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$		

Tabla 10.2: Definiciones de las relaciones topológicas relativas por medio de la matriz de proposiciones de orden

### 10.2.1. El problema de navegación topológica cualitativa

La HSCT encuentra una solución topológica cualitativa a un problema de navegación topológica cualitativa con la siguiente definición  $((X, \mathcal{T}), O, e_1, F, R, \Omega, \Lambda)$  donde  $R$  es el conjunto de las RRTTRR  $S_{13}$ . Pero la HSCT también puede ser usada para establecer relaciones topológicas clásicas donde  $R$  es el conjunto de relaciones  $R_8$ . Esto es posible porque cada una de las RRTTRR está contenida solo en una relación topológica clásica y las relaciones topológicas clásicas de  $R_8$  son disjuntas.

$$\begin{aligned} \forall s \in S_{13} \quad \exists! r \in R_8, \quad s \subseteq r \\ \forall r, r' \in R_8, \quad r \neq r' \quad \rightarrow \quad r \cap r' = \emptyset \end{aligned}$$

Así, la HSCT puede resolver el establecer relaciones topológicas clásicas si se transforma la relación topológica clásica objetivo por una RTR que cumpla que está contenida en la relación topológica clásica objetivo. La transformación de la relación topológica clásica a la RTR se puede llevar a cabo de diferentes maneras. Por ejemplo, se puede elegir aleatoriamente, se puede elegir siempre de una manera predeterminada, o elegir mediante algún criterio como la distancia.

### 10.2.2. Generalización a n-dimensiones

La heurística que se ha presentado en el apartado anterior permite encontrar soluciones a problemas del tipo  $((\mathbb{Z}, \mathcal{T}), O, e_1, F, R, \Omega, \Lambda)$  siendo  $R$  el conjunto de relaciones topológicas relativas; pero un algoritmo para toma de decisiones de navegación en espacios de una dimensión es poco útil para el mundo real que consta al menos de tres. Así, la cuestión siguiente en la investigación ha sido la generalización de la heurística a dimensión  $n$ . Por ello, en este apartado se presenta la investigación que se ha realizado para la aplicación de los GGLRT a los problemas  $((\mathbb{Z}^n, \mathcal{T}), O, e_1, F, R, \Omega, \Lambda)$ , aunque es evidente que el interés real surge para  $n = 2$ , o  $n = 3$ .

La primera cuestión sobre la generalización es sobre la creación de un espacio topológico  $n$ -dimensional. Para ello, se pueden usar varios resultados de topología que permiten construir el espacio topológico  $\mathbb{Z}^n$ . Se sabe que dada una topología sobre  $X$  y otra sobre  $Y$  hay una manera canónica de crear una topología sobre el producto cartesiano  $X \times Y$ , la denominada topología producto<sup>2</sup>. Así, a partir de la base  $\mathcal{B} = \{(x_1, x_2) : x_1 < x_2 \quad x_1, x_2 \in \mathbb{Z}\}$  que genera la topología del orden sobre  $\mathbb{Z}$  se puede construir la topología producto sobre  $\mathbb{Z} \times \mathbb{Z}$ . La definición de la topología producto sobre  $X \times Y$  se puede aplicar de igual manera a  $X_1 \times X_2 \times \dots \times X_n$ . Por lo tanto, no hay problema en construir una generalización a dimensión  $n$  del espacio topológico.

La siguiente cuestión es cuáles son las relaciones topológicas en un espacio  $X_1 \times X_2 \times \dots \times X_n$ . Debido a que el espacio es construido mediante el producto cartesiano

<sup>2</sup>Ver anexo D

la relación topológica  $n$ -dimensional entre dos objetos será definida por una tupla de  $n$  componentes siendo el componente  $k$  la relación topológica que se da en la dimensión  $k$ -ésimo entre los dos objetos. Así :

$$\forall A, B \in X_1 \times \cdots \times X_n \quad Ar_x B \Leftrightarrow A \langle r_{x_1}, r_{x_2}, \dots, r_{x_n} \rangle B$$

$r_{x_k}$  es la relación que se da entre  $A$  y  $B$  en la dimensión  $k$ .

Así, para expresar la relación espacial “sobre” y “dentro de” en un espacio tridimensional se representan de la siguiente manera:

$$o_x \langle \text{contiene}, \text{contiene}, \text{pegados} - 1 \rangle o_y \quad \equiv \quad o_x \text{ sobre } o_y$$

$$o_x \langle \text{contiene}, \text{contiene}, \text{contiene} \rangle o_y \quad \equiv \quad o_y \text{ dentro de } o_x$$

Por lo tanto, la generalización a  $n$  dimensiones de la heurística HSCT consiste en aplicarla sucesivamente sobre cada dimensión. Aunque esta generalización tiene varias restricciones respecto al mundo real.

Uno de los problemas que aparece al cambiar de  $\mathbb{Z}$  a  $\mathbb{Z}^n$  es el del contorno de los objetos. En  $\mathbb{Z}$  los objetos tienen un contorno que no varía respecto a los puntos de referencia que se toman para calcular la relación topológica relativa, ya que el contorno es sólo un punto en cada extremo del intervalo. En  $\mathbb{Z}^n$  puede tener cualquier tipo de contorno, de manera que los puntos del intervalo pueden tener valores que no se corresponden con los valores de otros puntos del contorno. Así, para la generalización es necesario que los lados de cada objeto tengan ángulos rectos entre sí, de manera que en espacios de 2 dimensiones los objetos sean rectángulos, en espacios de 3 dimensiones los objetos sean cuboides, etc. Eso asegura que se pueda modificar en otra dimensión la correspondiente relación topológica sin que esto afecte a la anterior. Es decir, se cumple el principio de optimalidad, y la combinación de una solución para cada dimensión es una solución del problema completo.

Por lo tanto, la siguiente condición se necesita para que exista el principio de optimalidad para los problemas  $((\mathbb{Z}^n, \mathcal{T}), O, e_1, R, \Theta, \Lambda)$ , y permite que se pueda aplicar la HSCT a cada dimensión para obtener la solución.

$$\forall (o_j, A_i) \quad A_i \in \mathcal{B}_{\mathbb{Z} \times \dots \times \mathbb{Z}} = \{((x'_1, x'_2), \dots, (x_1^n, x_2^n)) : x_1, x_2, \dots, x_1^n, x_2^n \in \mathbb{Z}\}$$

Al cumplirse la anterior condición en el espacio  $\mathbb{Z}^2$  los objetos van a tener la forma  $((x_1, x_2), (y_1, y_2))$  que se visualiza en la figura 10.2.

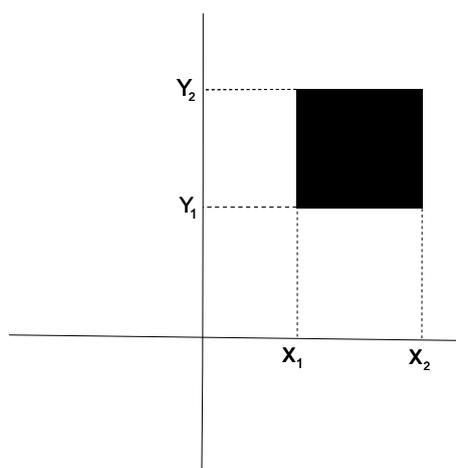


Figura 10.2: Representación gráfica de un elemento  $((x_1, x_2), (y_1, y_2))$  de la base  $\mathcal{B}_X \times \mathcal{B}_Y$

Pero la posible utilidad de los GGLRT sobre un espacio de dimensión  $n > 1$  construido con la topológica producto no está sólo en relación con la forma de los objetos, ya que también las acciones pueden actuar en varias dimensiones. Por eso, para que el principio de optimalidad se cumpla también se ha de tener en cuenta sobre cuantas dimensiones actúan las acciones. Si las acciones actúan sobre los valores de más de una dimensión entonces no se puede tratar la navegación por cada dimensión como un problema independiente. Así, para que la HSTC pueda ser aplicada sobre cada dimensión, además de la condición de que los lados de cada objeto formen ángulos rectos entre si, cada función  $f_i$  debe modificar sólo los valores de una dimensión.

### 10.3. Extensiones de la HSCT: relaciones borrosas y de cercado

En La sección anterior se ha presentado el resultado del que ha partido el programa RMMNT, la HSCT. La HSCT es una heurística para resolver una versión simple del problema de navegación basada en relaciones espaciales. Así, la heurística permite tomar decisiones sobre navegación local para establecer relaciones espaciales. Ese primer resultado permite estudiar la creación de variantes de la heurística que permitan abordar situaciones espaciales más complejas. En esta sección se presentan dos líneas de investigación que se han comenzado para extender el alcance de la HSCT. Una línea consiste en añadir relaciones topológicas de cercado para describir situaciones espaciales no contempladas por las relaciones topológicas relativas. La otra línea aborda el uso de modificadores borrosos con el objetivo de controlar la velocidad de navegación.

### 10.3.1. Extensión de HSCT añadiendo relaciones de cercado

Los modelos clásicos de razonamiento topológico contienen 8 relaciones binarias diferentes: disjuntos, pegados, solapados, cubierto, cubre, contenido, iguales, contiene, [484]. En esos modelos hay un importante tipo de relaciones topológicas que no es representado, las relaciones de cercamiento. Las relaciones de cercamiento no son las de “rodea a”. Mientras dos regiones espaciales, o una región espacial y un objeto, pueden compartir el mismo espacio, dos objetos no pueden compartir el mismo espacio y mantener sus integridades como objetos en un entorno físico clásico. Así, estas relaciones son muy importantes para describir situaciones como por ejemplo “el castillo está rodeado por un foso”. Por esta razón, es importante investigar modelos de razonamiento que consideren las relaciones de cercado. También son interesantes para describir situaciones en las cuales las regiones u objetos espaciales tienen agujeros. En el campo de los sistemas de información geográfica se ha estudiado la representación de regiones topológicas con agujeros [141, 510, 362, 250, 265]. Sin embargo, en el campo de la robótica no se han realizado investigaciones sobre este tema antes de que esta investigación se llevara a cabo. Así, se ha llevado a cabo un análisis general sobre cómo incorporar la descripción de objetos con agujeros a la HSCT. En el análisis se ha considerado el espacio  $(\mathbb{Z}, \mathcal{T})$  siendo  $\mathcal{T}$  la topología del orden y las relaciones espaciales entre objetos en ese espacio.

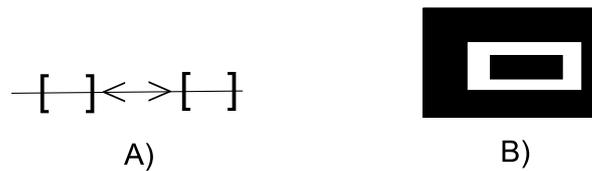


Figura 10.3: Relaciones de cercado. A) Una relación de cercado entre dos objetos en un espacio unidimensional. B) Una relación de cercado entre dos objetos en un espacio bidimensional.

Una vez se logra un algoritmo para un espacio unidimensional, se puede formular una generalización para un espacio  $n$ -dimensional como se ha visto en el apartado anterior. Durante el análisis que se llevó a cabo, se presupuso que los objetos pueden tener máximo un agujero. Por lo tanto, hay dos parámetros que deben ser evaluados para determinar el GRLT cuando poseen relaciones de cercado para aplicar la heurística: el número de objetos que tienen agujeros y las relaciones de tamaño entre los objetos con agujero. Las situaciones de dos objetos con como máximo un agujero son las siguientes:

- Ninguno de los objetos tiene agujeros.
- Solo uno de los objetos tiene un agujero.
- Ambos objetos tienen un agujero.

En la situación de que ninguno de los objetos tiene agujeros, hay 6 subcasos relativos al tamaño de los objetos como se mostró en el apartado anterior. En la situación del objeto con agujero y el objeto sin agujero, es importante notar que un objeto con un agujero puede ser descrito como una tupla de tres zonas de espacio, que se puede describir como  $O = (O_1, H_O, O_2)$ . Dado que las posibles relaciones de tamaño entre cada una de las tres partes del objeto con un agujero y el objeto sin agujero son 6, el número de subcasos a contemplar entre un objeto que tiene un agujero y un objeto sin un agujero son  $6^3 = 216$  subcasos posibles. Cada uno de esos subcasos implica un GRLT para la toma de decisiones. En la situación en que ambos objetos tienen agujero, el número de subcasos es  $6^9 = 10,077,696$ . Por razones de espacio, no se puede no se puede respresentar una tabla con el GRLT para cada posible subcaso. Sin embargo, ahora se sabe de la compleja tarea de programación que supondría desarrollar un programa para la toma de deciones de navegación para situaciones en el que los objetos o regiones poseen un agujero. Sin embargo, el tiempo de ejecución requerido para tomar una decisión no se incrementaría respecto al de HSCT porque consiste en recuperar un objeto de una tabla.

Respecto al método para calcular una relación topológica relativa entre dos objetos, se ha de construir una nueva matriz de proposiciones de orden para que pueda determinar todas las posibles relaciones espaciales. Para la situación en la que el agente  $X$  no posee un agujero y el objeto de referencia tiene uno, la matriz de proposiciones de orden es la siguiente:

$$P^{\leq}(X, Y) = \begin{pmatrix} \text{mín}(X) < \text{mín}(Y) & \text{mín}(X) < \text{máx}(Y) \\ \text{máx}(X) < \text{mín}(Y) & \text{máx}(X) < \text{máx}(Y) \\ \text{mín}(X) = \text{mín}(Y) & \text{mín}(X) - 1 = \text{máx}(Y) \\ \text{máx}(X) + 1 = \text{mín}(Y) & \text{máx}(X) = \text{máx}(Y) \\ \text{mín}(X) < \text{mín}(H_Y) & \text{mín}(X) < \text{máx}(H_Y) \\ \text{máx}(X) < \text{mín}(H_Y) & \text{máx}(X) < \text{máx}(H_Y) \\ \text{mín}(X) = \text{mín}(H_Y) & \text{mín}(X) - 1 = \text{máx}(H_Y) \\ \text{máx}(X) + 1 = \text{mín}(H_Y) & \text{máx}(X) = \text{máx}(H_Y) \end{pmatrix}$$

En el caso en el que el agente  $X$  y el objeto de referencia tiene un agujero, la matriz de proposiciones de orden es:

$$P^{\leq}(X, Y) = \left( \begin{array}{ll} \text{mín}(X) < \text{mín}(Y) & \text{mín}(X) < \text{máx}(Y) \\ \text{máx}(X) < \text{mín}(Y) & \text{máx}(X) < \text{máx}(Y) \\ \text{mín}(X) = \text{mín}(Y) & \text{mín}(X) - 1 = \text{máx}(Y) \\ \text{máx}(X) + 1 = \text{mín}(Y) & \text{máx}(X) = \text{máx}(Y) \\ \text{mín}(X) < \text{mín}(H_Y) & \text{mín}(X) < \text{máx}(H_Y) \\ \text{máx}(X) < \text{mín}(H_Y) & \text{máx}(X) < \text{máx}(H_Y) \\ \text{mín}(X) = \text{mín}(H_Y) & \text{mín}(X) - 1 = \text{máx}(H_Y) \\ \text{máx}(X) + 1 = \text{mín}(H_Y) & \text{máx}(X) = \text{máx}(H_Y) \\ \text{mín}(H_X) < \text{mín}(Y) & \text{mín}(H_X) < \text{máx}(Y) \\ \text{máx}(H_X) < \text{mín}(Y) & \text{máx}(H_X) < \text{máx}(Y) \\ \text{mín}(H_X) = \text{mín}(Y) & \text{mín}(H_X) - 1 = \text{máx}(Y) \\ \text{máx}(H_X) + 1 = \text{mín}(Y) & \text{máx}(H_X) = \text{máx}(Y) \\ \text{mín}(H_X) < \text{mín}(H_Y) & \text{mín}(H_X) < \text{máx}(H_Y) \\ \text{máx}(H_X) < \text{mín}(H_Y) & \text{máx}(H_X) < \text{máx}(H_Y) \\ \text{mín}(H_X) = \text{mín}(H_Y) & \text{mín}(H_X) - 1 = \text{máx}(H_Y) \\ \text{máx}(H_X) + 1 = \text{mín}(H_Y) & \text{máx}(H_X) = \text{máx}(H_Y) \end{array} \right)$$

### 10.3.2. Extension de la HCST mediante modificadores borrosos

La formulación inicial del problema de navegación mediante relaciones espaciales para el que se ha desarrollado la HCST considera únicamente acciones en las cuales llevar a cabo una acción supone siempre que se avanza una unidad espacial. Una formulación más general del problema de navegación de relaciones espaciales supone considerar acciones donde el número de unidades que se desplaza el agente sea variable. Puesto que el espacio que se considera para los agentes es discreto, el número de unidades que se desplaza un agente puede ser considerado su velocidad. Así, se puede plantear no sólo el objetivo de establecer una relación espacial sino hacerlo en el menor tiempo posible sin movimientos bruscos. La solución en términos generales es que el agente se mueva a la mayor velocidad posible cuando está lejos de la relación objetivo y que vaya reduciendo su velocidad según se acerque a la relación objetivo. El problema es que las relaciones topológicas saltan bruscamente de una situación a otra. Por ejemplo, la relación topológica de *disjuntos* contiene situaciones con distancias muy lejanas y muy cercanas, y la siguiente relación topológica que es *tocando* implica distancia 0. La línea que se ha investigado para resolver este problema ha sido extender HCST añadiendo modificadores borrosos.

En los campos de sistemas de información geográfica y razonamiento automático se ha investigado la definición y el uso de relaciones topológicas borrosas [214, 560, 308, 58]. Sin embargo el concepto de relación topológica borrosa no se ha empleado en robótica

antes de la investigación que aquí se presenta. La idea clave que se ha investigado para controlar la velocidad ha sido extender la HTQS definiendo una relación borrosa complementaria,  $s_i^c$ , para cada una de las relaciones topológicas relativas  $s_i$ . La figura 10.4 muestra un ejemplo de una relación topológica y su complementaria en el marco de las relaciones borrosas.

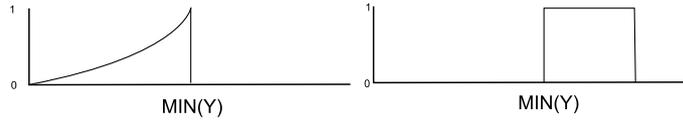


Figura 10.4: Izquierda. Representación del grado de verdad de la relación complementaria de *solapados-0*. Derecha. Representación del grado de verdad de la relación *solapados-0*.  $Y$  es el objeto objetivo.  $\text{Min}(Y)$  determina la distancia al punto más cercano de  $Y$  al agente.

El método que se ha desarrollado trabaja de la siguiente manera. Dada la RTR objetivo, se calcula el valor de verdad de su función complementaria, y se determina que la velocidad a usar por el agente es inversamente proporcional al valor de verdad de la relación complementaria. Así, la velocidad que el agente escogería vendría determinada por la siguiente expresión:

$$velocidad = \lceil (velocidad_{max} - 1) \cdot (1 - \mu_{s_i^c}(X, Y)) \rceil$$

donde  $velocidad_{max}$  es la velocidad máxima que puede desarrollar el agente,  $\mu_{s_i^c}$  la función de verdad de la relación  $s_i^c$ ,  $X$  es el espacio ocupado por el agente e  $Y$  el espacio ocupado por el objeto con el que se pretende establecer la relación  $s_i$ . La velocidad se considera que siempre será un número positivo entero porque se considera que el agente y el objeto están en un espacio discreto.

La cantidad resultante del producto es redondeada hacia arriba porque si la velocidad es entre 0 y 1 el resultado sería 1 lo que asegura que siempre se llegará al objetivo. Sin embargo, si la velocidad es redondeada hacia abajo el resultado es 0 cuando la velocidad es entre 0 y 1, lo que produciría que el agente nunca lograría su objetivo al estar detenido. Además, se podría definir un umbral de verdad para  $\mu$  con el objetivo de mantener más tiempo la máxima velocidad. Así, la velocidad quedaría definida por la siguiente función:

$$velocidad = \begin{cases} \lceil velocidad_{max} \cdot (1 - \mu_{s_i}(X, Y)) \rceil & \mu_{s_i}(X, Y) \geq u \\ velocidad_{max} & \mu_{s_i}(X, Y) < u \end{cases}$$

Dado el método anterior para seleccionar la velocidad, la siguiente cuestión es determinar las funciones  $\mu_{s_i}$ . Las propuestas que se han hecho hasta ahora de relaciones topológicas borrosas se han basado siempre en considerar un espacio topológico borroso o regiones borrosas. Sin embargo, en este caso no se tiene un espacio topológico borroso

sino un espacio topológico. Dado que en las RRTTR se distingue la dirección del movimiento para establecer la relación objetivo, es importante desde qué lado está partiendo el agente. Esto implica que la función de verdad debe trabajar con dos funciones, una para cada uno de los lados, y que se denotarán como:  $\mu_{s_i^c}^>$  y  $\mu_{s_i^c}^<$ . La HTQS borrosa calcula desde que lado está partiendo el agente hacia la relación objetivo mediante la verificación de qué relación de orden se mantiene entre  $n_{s_c}$  y  $n_{s_t}$ . Así, el orden de la relación es usado para seleccionar la función:

$$\mu_{s_i^c}(X, Y) = \begin{cases} \mu_{s_i^c}^>(X, Y) & n_{s_c} > n_{s_t} \\ \mu_{s_i^c}^<(X, Y) & n_{s_c} < n_{s_t} \end{cases}$$

Las funciones de verdad que se proponen para las relaciones complementarias  $s_i^c$  son las de la tabla 10.3.

$s_i^c$	$\mu_{s_i^c}^>(X, Y)$	$\mu_{s_i^c}^<(X, Y)$
$s_1^c$ (Casi disjuntos-0)	$\frac{1}{ (\max(X)+1)-\min(Y) +1}$	$\frac{1}{ (\max(X)+1)-\min(Y) +1}$
$s_2^c$ (Almost pegados-0)	$\frac{1}{ (\max(X)+2)-\min(Y) +1}$	$\frac{1}{ (\max(X)-\min(Y)) +1}$
$s_3^c$ (Casi solapados-0)	$\frac{1}{ (\max(X)+1)-\min(Y) +1}$	$\frac{1}{ (\min(X)-\min(Y)) +1}$
$s_4^c$ (Casi cubierto-0)	$\frac{1}{ (\min(X)+1)-\min(Y) +1}$	$\frac{1}{ (\min(X)-1)-\min(Y) +1}$
$s_5^c$ (Casi cubre-0)	$\frac{1}{ (\max(X)+1)-(\max(Y)) +1}$	$\frac{1}{ (\max(X)-1)-(\max(Y)) +1}$
$s_6^c$ (Casi contenido)	$\frac{1}{ \min(X)-\min(Y) +1}$	$\frac{1}{ \max(X)-\max(Y) +1}$
$s_7^c$ (Casi iguales)	$\frac{1}{ (\min(X)+1)-\min(Y) +1}$	$\frac{1}{ (\max(X)-1)-\max(Y) +1}$
$s_8^c$ (Casi contiene)	$\frac{1}{ \max(X)-\max(Y) +1}$	$\frac{1}{ \min(X)-\min(Y) +1}$
$s_9^c$ (Casi cubre-1)	$\frac{1}{ (\min(X)+1)-\min(Y) +1}$	$\frac{1}{ (\min(X)-1)-(\min(Y)) +1}$
$s_{10}^c$ (Casi cubierto-1)	$\frac{1}{ (\max(X)+1)-\max(Y) +1}$	$\frac{1}{ (\max(X)-1)-\max(Y) +1}$
$s_{11}^c$ (Casi solapados-1)	$\frac{1}{ \max(X)-\max(Y) +1}$	$\frac{1}{ (\min(X)+1)-\max(Y) +1}$
$s_{12}^c$ (Casi pegados-1)	$\frac{1}{ (\min(X)-\max(Y)) +1}$	$\frac{1}{ (\min(X)+2)-\max(Y) +1}$
$s_{13}^c$ (Casi disjuntos-1)	$\frac{1}{ (\min(X)-1)-\max(Y) +1}$	

Tabla 10.3: Funciones propuestas para las relaciones topológicas relativas complementarias borrosas.

A continuación se presenta un ejemplo numérico en el que la HSCT borrosa es usada para decidir la velocidad que debe usar un agente. Dado el espacio unidimensional  $\mathbb{Z}$  en el que un agente  $o_A$  tiene el objetivo de establecer una relación espacial con la región espacial  $o_R$ . Las zonas espaciales ocupadas inicialmente por  $A$  y  $R$  son  $o_A = [1, 6]$  y  $o_R = [20, 26]$ . El agente  $A$  tiene las tres siguientes acciones para desplazarse por el espacio con una velocidad variable:

Función	Etiqueta
$f_1$	$<$
$f_2$	$=$
$f_3$	$>$

Tabla 10.4: Tabla de medios y fines del ejemplo con las acciones etiquetadas.

$$\Lambda(o_A) = \{f_{(1,v)}|_x(x) = x + v, \quad f_2|_x(x) = x, \quad f_{(3,v)}|_x(x) = x - v\}$$

Así, las acciones de  $A$  son etiquetadas y quedan descritas en la tabla 10.4.

Puesto que tanto el espacio como el tiempo son discretos en este ejemplo, se entenderá por velocidad las unidades discretas que se desplaza el agente con cada decisión que toma. En este caso la velocidad máxima que puede desarrollar  $A$  será de 10 unidades. La meta del agente es lograr la relación *pegados-0* con la región espacial  $R$ . La HTQS borrosa calcula que el GLRT que debe ser usado es el número 2. El valor de  $n_{s_c}$  es 1, y el valor de  $n_{s_t}$  es 2, así que la relación de orden es  $<$ . Con esos datos, la HTQS borrosa selecciona la acción  $f_1$ . Dado que la relación que se debe de establecer es *pegados-0*, la relación topológica relativa complementaria es *Casi pegados-0*. Usando la relación de orden para seleccionar la función de verdad en la tabla 10.3 se determina que la que se usará es la siguiente:

$$\frac{1}{|(\max(X) + 2) - \min(Y)| + 1}$$

Por lo tanto, la velocidad que aplica el agente  $A$  al utilizar la HTQS borrosa en el instante inicial es:

$$v = \lceil (\text{velocidad}_{max} - 1) \cdot (1 - \mu_{s_c}^c(X, Y)) \rceil = \lceil 9 \cdot (1 - \frac{1}{|(6 + 2) - 20| + 1}) \rceil = \lceil 8, 3 \rceil = 9$$

Tras moverse las 9 unidades, el agente tiene que volver a tomar una decisión, en esta ocasión el espacio ocupado por el agente es  $o_A = [10, 15]$ . De nuevo, la HTQS borrosa selecciona la acción  $f_1$  y la misma relación topológica relativa complementaria. Pero en esta ocasión la velocidad elegida es

$$v = \lceil (\text{velocidad}_{max} - 1) \cdot (1 - \mu_{s_c}^c(X, Y)) \rceil = \lceil 9 \cdot (1 - \frac{1}{|(15 + 2) - 20| + 1}) \rceil = 7$$

De esta manera, la HTQS borrosa selecciona diferentes velocidades y logra un movimiento suave para establecer una relación topológica relativa. En este caso se ha

supuesto que el tiempo es discreto y que el agente se desplaza todas las unidades espaciales antes de que comience la nueva toma de decisiones. Si el agente tomara una nueva decisión antes de que se haya desplazado todas las unidades que se han fijado como velocidad, la curva de velocidad sería más suave aún.

## 10.4. Arquitectura de navegación cualitativa topológica

Hasta el momento, la mayoría de algoritmos usados para la navegación han tenido el propósito de dirigir al robot a un punto del espacio. Sin embargo, como ya se ha comentado, los seres humanos se comunican tareas mediante la especificación de relaciones espaciales entre elementos o lugares. Además, los entornos en los cuales los humanos desarrollan sus actividades son extremadamente dinámicos. La única opción que permite la navegación exitosa en entornos dinámicos y desconocidos es la toma de decisiones en tiempo real. Por lo tanto, para que un robot sea capaz de colaborar con los seres humanos debe ser capaz de tomar decisiones basadas en información local registrada por sus sensores e interpretar y expresar relaciones espaciales. Además, cuando a una persona se le indica una tarea para realizar de manera continuada en un entorno, esa tarea se le comunica dándole unas metas genéricas, de manera que la persona no necesite ser supervisada constantemente. Así, dos nuevos problemas aparecen cuando se quieren crear robots multifuncionales: cómo navegar en un entorno dinámico y desconocido usando relaciones espaciales para realizar tareas y cómo lograr lo anterior sin una supervisión. En esta sección, se presenta una nueva arquitectura para abordar los problemas anteriores, denominada *arquitectura de navegación cualitativa topológica (ANCT) 1.0*. En las anteriores secciones se ha presentado la HSCT y cómo es capaz de identificar y establecer relaciones espaciales. Sin embargo, la HSCT solo permite establecer una relación espacial con un objeto específico. En cambio, la navegación requiere una secuencia temporal de metas con diferentes objetos. La arquitectura que se presenta genera automáticamente nuevas metas y las resuelve usando HSCT. Así, la nueva arquitectura permite realizar navegación autónoma en entornos abiertos dinámicos o desconocidos. Esta sección se organiza de la siguiente manera. El apartado 10.4.1 presenta y explica todos los conceptos teóricos que usa la ANCT 1.0. La sección 10.4.2 explica cómo se implementa la ANCT 1.0 y las pruebas experimentales que se han hecho con ella.

### 10.4.1. Fundamentos de la arquitectura de navegación cualitativa topológica

Este apartado presenta los conceptos relacionados con la generación de comportamientos que se usan en la ANCT 1.0., y cómo se han empleado para diseñar la ANCT 1.0. La ANCT 1.0 tiene cuatro características principales: reactividad de segundo orden, proactividad, autonomía basada en tareas topológicas y un sistema de referencia egocéntrico. La primera parte de este apartado presenta el mecanismo que genera el comportamiento proactivo y reactivo que incorpora la ANCT 1.0. La segunda parte demostrará cómo la ANCT 1.0 incorpora tareas genéricas y específicas para evitar la necesidad de supervisión constante. El tercer apartado presenta una descripción general de los procesos y subprocesos de la arquitectura. La última parte explicará en detalle la implementación de dos tareas topológicas que permiten al agente llevar a cabo comportamientos de exploración y de colocación para alimentarse en un micromundo virtual.

#### 10.4.1.1. Comportamientos reactivos y proactivos

En entornos que son desconocidos o cuya dinámica es compleja, no es posible usar un algoritmo de planificación desconectado para lograr una meta. Las investigaciones en robótica han mostrado que las arquitecturas con métodos reactivos son los más efectivos para la navegación en entornos desconocidos [81, 322, 255, 17, 496]. El comportamiento reactivo implica la existencia de respuestas instantáneas de un sistema asociadas a la llegada de estímulos específicos al sistema. El comportamiento reactivo da buenos resultados cuando el sistema necesita responder en tiempo real. Ese éxito llevó a Rodney Brooks a proponer la hipótesis del anclaje físico [84]. La hipótesis del anclaje físico afirma que un sistema inteligente necesita que su representación esté anclada en el mundo físico. Brooks afirmó:

*“The key observation is that the world is its own best model. It is always exactly up to date. It always contains every detail there is to be known. The trick is to sense the world appropriately and often enough.”* [84] p.5.

Los seres humanos reaccionan a sensaciones o percepciones. Mientras una sensación es una sencilla estructura de información, una percepción es una estructura de información compleja. Así, cuando un ser humano se quema en una mano tras colocarla sobre algún objeto con alta temperatura, la reacción de la persona es retirar inmediatamente la mano debido a la señal dolorosa que se produce. Cuando un objeto se acerca rápidamente hacia la cara de una persona, los ojos de la persona se cerrarán incluso moverá su cabeza, en este caso se está reaccionando a la percepción de que un objeto se está aproximando. En esta situación, la reacción es más compleja porque la persona está

detectando un objeto que se está moviendo hacia la cara de la persona. El objetivo del programa RMMNT es para aplicar nociones topológicas. Inevitablemente, ese objetivo implica que la información del entorno tiene que estar organizada en objetos y relaciones entre ellos en el sistema. La reactividad que se debe investigar en el programa RMMNT es una respuesta a las relaciones entre agentes y objetos en el entorno. Esto difiere de los algoritmos reactivos desarrollados hasta ahora, ya que estos usan señales directas capturadas por los sensores que son valores cuantitativos calculados a partir de los sensores como parachoques o infrarrojos. Utilizando la distinción entre los dos tipos de reactividad, la reactividad investigada por el programa RMMNT es reactividad de segundo orden, mientras que la reactividad desarrollada hasta ahora en los algoritmos, como Bug, es denominada reactividad de primer orden<sup>3</sup>. Más adelante se mostrará en detalle la ANCT 1.0 reacciona a relaciones que son encontradas en su entorno mediante la HSCT.

Entre las muchas habilidades necesitadas por un robot móvil para ser capaz de trabajar en un entorno desconocido o dinámico, dos de las más importantes son: (1) la capacidad para navegar evitando obstáculos y (2) la capacidad para explorar comportamiento. Respecto a la capacidad de un robot de evitar obstáculos usando relaciones topológicas, una posible estrategia es usar la HSCT de manera que el robot sistemáticamente evite todas las relaciones espaciales que implican su colisión con otros objetos. Sin embargo, esta estrategia puede privar al robot de su capacidad de explorar autónomamente porque si el robot se encuentra en una situación en la que evita las colisiones no se moverá de ella porque ese es el objetivo de la estrategia. Otro problema con esa estrategia es que una acumulación de objetos en el espacio requeriría al método de toma de decisiones un incremento en el coste computacional porque tendría que evaluar cada relación espacial con cada objeto. En los robots, el comportamiento exploratorio resulta de comportamientos proactivos; por lo tanto, generar este tipo de comportamiento es esencial. La estrategia que se ha escogido en la investigación es generar proactividad por medio de la reactividad. La idea clave para generar comportamiento proactivo a partir de reactividad es que el agente cree en su representación del espacio objetos virtuales que no tendrán existencia física. Los objetos virtuales provocarán que el agente reaccione a ellos, pero para un observador externo el agente se estará comportando de manera proactiva. En su entorno, un agente va a encontrar una gran cantidad de espacio vacío. Los objetos virtuales son objetos que tienen regiones de espacio vacío asignadas. Los objetos virtuales permiten a los agentes tratar de establecer relaciones espaciales con las regiones espaciales de espacio vacío asignadas a ellos. Por lo tanto, en la representación del espacio de la ANCT, existen dos tipos de clases de objetos: reales y virtuales. La ANCT 1.0 es tanto reactiva como proactiva. Los objetos virtuales

---

<sup>3</sup>Esta nomenclatura se ha tomado de la lógica donde se tiene lógica de primer orden y segundo orden. La lógica de primer orden tiene solo cuantificación sobre los elementos mientras que la lógica de segundo orden tiene cuantificación sobre las relaciones.

que se usan para la navegación son denominados *puertas*. La idea es que los agentes tendrán como objetivo pasar a través de las puertas. La generación de las puertas en la representación del espacio se puede realizar mediante diferentes estrategias. Las tres estrategias que se proponen (que se pueden ver en la figura 10.5) son las siguientes:

- Generar puertas en los límites de los sensores.
- Generar puertas al final del primer obstáculo.
- Generar puertas al principio del primer obstáculo.

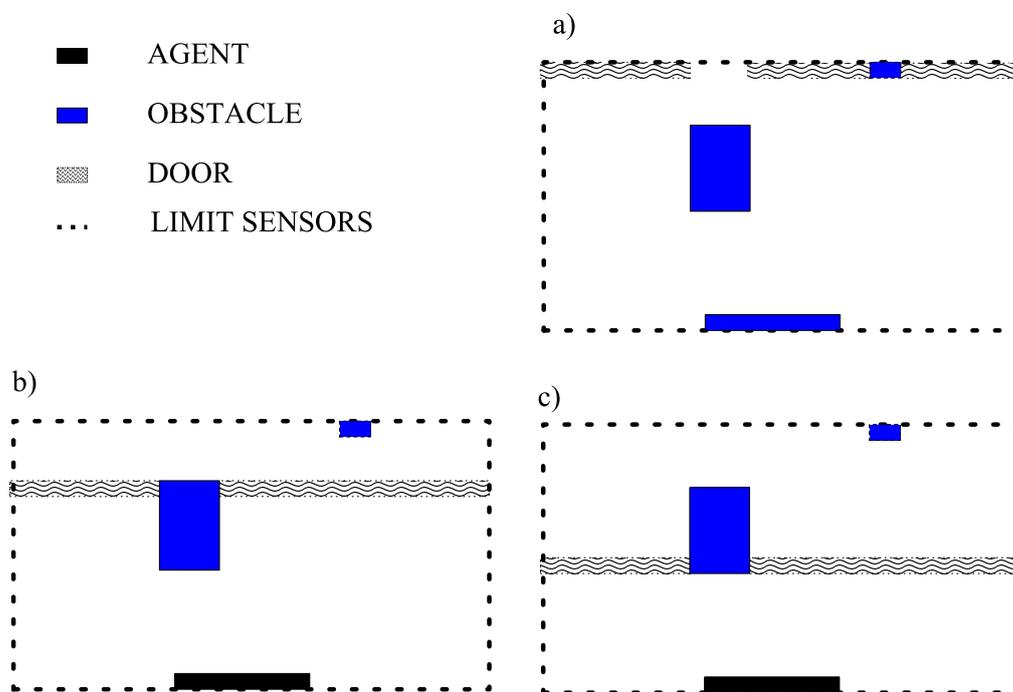


Figura 10.5: Ejemplo gráfico de las diferentes estrategias para generar puertas en un entorno por un agente. A) Generar puertas en los límites de los sensores. B) Generar puertas al final del primer obstáculo. C) Generar puertas al principio del primer obstáculo.

#### 10.4.1.2. Autonomía mediante tareas topológicas

Como se ha mencionado, una de las metas de la ANCT es proporcionar al agente un grado de autonomía, lo que implica que el agente sea capaz de autoseleccionar objetivos específicos. En las anteriores secciones la forma para representar un objetivo específico

en el espacio para un agente se ha formulado como el problema de establecimiento de relaciones topológicas, lo que determina a un agente el objetivo de establecer una relación espacial concreta con un objetivo particular. La idea que se ha explorado durante la investigación para dotar de autonomía a un agente ha sido que el agente genere autónomamente problemas de establecimiento de relaciones topológicas. Para lograr ese objetivo, es necesario que la arquitectura posea objetivos generales y capture información del entorno. Así, el método desarrollado para llevar a cabo esa idea hace uso de tres tipos de elementos. El primer elemento, que será denominado *tarea topológica* es un conjunto de fórmulas que describen una tarea de una manera general. Una tarea topológica no hace referencia a objetos específicos sino a objetos de un tipo particular. La idea de una tarea topológica es que contiene la información que dice al sistema qué debería hacer cuando encuentra un objeto que pertenece a una clase particular de objetos. Por ejemplo, si encuentras una cartera en el suelo, recógela. Si los objetos de la tarea son virtuales, se empleará el término *tarea topológica virtual*; y si los objetos son reales se empleará el término *tarea topológica real*. El segundo elemento, que se denominará *diagrama topológico abierto local*, es un conjunto de fórmulas describiendo el entorno local del agente. El último elemento es una regla de inferencia que genera un problema de establecimiento de relaciones topológicas mediante los dos elementos anteriores. Dado que es necesario referenciar diferentes dominios, el lenguaje formal para las fórmulas que se ha escogido es el de la lógica multivariada [335]. A continuación, se dan las definiciones formales de los elementos que se han mencionado:

Una tarea topológica es un conjunto finito de sentencias condicionales <sup>4</sup>, denotado por  $T$ .

$$T = \{\varphi_1, \varphi_2, \dots, \varphi_c\}$$

Cada sentencia  $\varphi_x$ , a la que se dominará *problema topológico general*, cumple el siguiente esquema:

$$\forall o, \vec{p}_1, \dots, \vec{p}_k \quad \left( \phi(o, \vec{p}_1, \dots, \vec{p}_k) \wedge size^d(\mathbf{o}_A, o) = v \wedge \neg r_x^d(\mathbf{o}_A, o) \right) \rightarrow r_x^d(\mathbf{o}_A, o)$$

$$o \in O, \quad \vec{p}_1 \in P_{\mu(1)}, \dots, \vec{p}_k \in P_{\mu(k)} \quad \text{donde} \quad \mu(k) \in \{1, \dots, n\}$$

Cada elemento del conjunto  $O$  referencia a un objeto en el entorno del agente. La constante  $\mathbf{o}_A$  pertenece a  $O$  e identifica al agente que posee la ANCT. La constante  $n$  es el número de propiedades de los objetos en el entorno. Así, la fórmula  $\phi(o, \vec{p}_1, \dots, \vec{p}_k)$  fija un patrón que define una clase de objetos. Nótese que las variables de la fórmula que están cuantificadas universalmente son las propiedades que no son relevantes para definir el patrón, porque todos los valores que tomen la variable hacen verdad la fórmula,

<sup>4</sup>Recuerdese que, en lógica, una sentencia es una fórmula que no tiene variables libres.

y los que son relevantes para definir el patrón son fijados usando constantes dentro de la fórmula  $\phi$ .

La cláusula  $size^d(\mathbf{o}_A, o) = \mathbf{v}$  determina la relación de tamaño entre el agente y el objeto en la dimensión  $d$ . La última cláusula del antecedente  $\neg r_x^d(\mathbf{o}_A, o)$  significa que la relación espacial en la cual están el objeto y el agente en la dimensión  $d$  debe ser evitada. El consecuente del condicional,  $r_x^d(\mathbf{o}_A, o)$ , expresa una relación espacial en la dimensión  $d$  entre el agente y el objeto que el agente debe establecer. Esta cláusula del consecuente es denominada la meta topológica general.

El siguiente concepto es el *diagrama topológico abierto local* (DTAL). En teoría de modelo, un diagrama abierto se refiere al conjunto de sentencias atómicas y negaciones de sentencias atómicas de un lenguaje que son una enumeración completa de los elementos de la estructura y que son verdad en la estructura. En el caso de los robots, no se puede obtener una descripción completa del entorno dado que las capacidades de los sensores son limitadas. Así, el uso de la palabra “local” es para reflejar que está restringido a los límites de los sensores. Basado en lo anterior, el *diagrama topológico abierto local del estado  $e_j$  y el agente  $\mathbf{o}_A$* , que se denota como  $DTAL(e_j, \mathbf{o}_A)$ , se define como un conjunto de sentencias que describen topológicamente las relaciones entre el robot y los objetos que el robot puede capturar del entorno usando sus sensores.  $LOTD(e_j, \mathbf{o}_A)$  puede ser dividido en dos conjuntos: el *diagrama local de objetos* (DLO) y el *diagrama relacional topológico local* (DRTL) del estado del entorno  $e_j$  y el agente  $\mathbf{o}_A$ . Es decir:

$$DTAL(e_j, \mathbf{o}_A) = DLO(e_j, \mathbf{o}_A) \cup DRTL(e_j, \mathbf{o}_A)$$

El DLO contiene sentencias que son descripciones de objetos, y representa esos objetos que el agente registra con sus sensores en un instante concreto. Cada frase del DLO tiene la siguiente estructura:

$$DLO(e_j, \mathbf{o}_A) = \{\psi'_i\} \text{ donde } \psi'_i \equiv (\mathbf{o}_{B_i}, \vec{p}_1, \dots, \vec{p}_n)$$

Las sentencias del DRTL representan las relaciones topológicas que el agente tiene con los objetos que detecta en la dimensión  $d$ . Cada una de las frases tendrá la siguiente estructura:

$$r_x^d(\mathbf{o}_A, \mathbf{o}_{B_i})$$

La relación  $r_x^d$  puede ser una relación topológica clásica o una relación topológica relativa. En la sección se mostró que la HSCT puede establecer una relación topológica clásica transformándola en una relación topológica relativa 10.2.

El tercer elemento, como se ha mencionado, es una regla de inferencia. La regla de inferencia es el mecanismo por el cual el agente puede autogenerar objetivos. La regla de inferencia empareja metas generales con la descripción topológica del entorno para generar metas específicas. Específicamente, la regla de inferencia es la regla *Modus Ponens*. Así, usando los dos elementos anteriores resulta la siguiente definición:

$$\frac{\forall o, \vec{p}_1, \dots, \vec{p}_k \left( \psi(o, \vec{p}_1, \dots, \vec{p}_k) \wedge size^d(\mathbf{o}_A, o) = v \wedge \neg\theta(o) \right) \rightarrow \theta(o)}{(\mathbf{o}_{B_i}, \vec{p}_1, \dots, \vec{p}_n) \wedge size(\mathbf{o}_A, \mathbf{o}_{B_i}) = v \wedge r_x^d(\mathbf{o}_A, \mathbf{o}_{B_i})} \theta([o/\mathbf{o}_{B_i}])$$

$$\psi([o/\mathbf{o}], [\vec{p}_1/\vec{p}_1], \dots, [\vec{p}_k/\vec{p}_k]) = (\mathbf{o}, \vec{p}_1, \dots, \vec{p}_n) \quad k \leq n$$

$$\theta([o/\mathbf{o}]) \neq r_x^d(\mathbf{o}_A, \mathbf{o}_{B_i}), \quad r_x^d(\mathbf{o}_A, \mathbf{o}_{B_i}) \in LTRD(e_j, o_A)$$

La sentencias  $\theta([o/\mathbf{o}])$  se denominará *meta topológica específica*. La fórmula  $\theta([o/\mathbf{o}])$  generada por la regla de inferencia ha reemplazado la variable  $o$  de  $\theta(o)$  por la constante  $\mathbf{o}_{B_i}$ . El tipo de objeto es determinado por un patrón en las características del objeto. En la segunda premisa de la regla, la primera cláusula es una fórmula del DLO, que hace el proceso de emparejamiento más rápido para la implementación de la arquitectura.

### 10.4.1.3. Los procesos de ANCT 1.0

La ANCT 1.0 está constituida básicamente por dos procesos generales. Uno de los procesos se ocupa de la generación de un objetivo particular en base a las tareas topológicas que tiene almacenadas y la información que ha registrado de su entorno. Este proceso es denominado *proceso autónomo de selección de objetivos* (PASO). El otro proceso lleva a cabo la toma de decisiones necesarias para lograr la meta específica que el PASO establece. Este proceso corresponde a una implementación de la HSCT. Estos dos procesos generales están relacionados de una manera secuencial. Además, cada uno de ellos está compuesto de diferentes subprocesos. El PASO está concretamente compuesto de tres subprocesos diferentes. El primer subproceso, denominado el *proceso de análisis del entorno* (PAE), analiza información del entorno para generar dos listas: objetos reales y objetos virtuales. Estas dos listas de objetos corresponden al DTAL (explicado en el apartado anterior 10.4.1.2). El segundo subproceso toma la lista de objetos generada por el PAE para buscar en ella si hay un objeto con el que haya asociada alguna tarea topológica, y si lo hay, entonces genera una meta topológica específica. El conocimiento que está almacenado mediante las tareas topológicas es fundamental para que la arquitectura sea exitosa llevando a cabo las tareas que se le requiere. Así, el siguiente apartado se centrará en explicar tareas topológicas. Este segundo subproceso es llevado a cabo por el motor de inferencia, que se describirá también en un apartado posterior. El tercer subproceso permite la modificación y la aceptación de metas en la base de tareas topológicas. Este último proceso es opcional y no ha sido programado en la implementación de la ANCT 1.0 en los experimentos. Las relaciones entre los distintos procesos y subprocesos de la ANCT 1.0 se muestran en la figura 10.6.

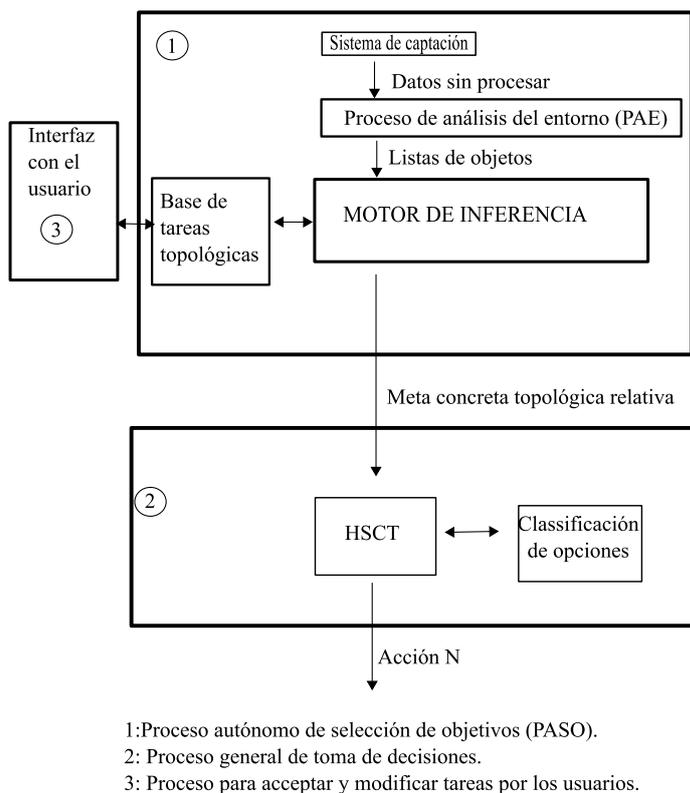


Figura 10.6: Diagrama cómo se relacionan los procesos y subprocesos de la ANCT 1.0

Se debe notar que el PASO genera un problema de establecimiento de relaciones topológicas y que HSCT solo genera una acción porque después de que HSCT devuelve una acción, el PASO se ejecuta de nuevo y así un nuevo problema de establecimiento de relaciones topológicas es generado. Si el entorno no cambia, el problema es el mismo y HSCT continúa resolviéndolo; si el entorno cambia, entonces el problema de establecimiento de relaciones topológicas es diferente. Esto permite que la ANCT 1.0 pueda afrontar entornos dinámicos y desconocidos.

#### 10.4.1.4. La implementación de tareas topológicas en la ANCT 1.0

Las tareas topológicas no solo proporcionan autonomía a los agentes, sino que además permiten el almacenamiento de información crítica para el funcionamiento propio de los agentes o robots. En este apartado se va a explicar detenidamente la implementación de dos tareas topológicas. Un hecho que se ha de tener en cuenta a la hora de implementar tareas topológicas es que el robot intentará lograr realizar las tareas que sean indicadas por el programador. Por lo tanto, el programador no debería nunca incluir una tarea que es físicamente imposible.

La implementación de tareas topológicas está íntimamente relacionada con el sistema de captación del entorno que el agente tiene incorporado. Las dos tareas topológicas que van a ser explicadas son implementadas de la manera en la que lo son porque la proyección del sistema visual de un agente del micromundo es ortográfica. Si la proyección del sistema visual no fuera ortográfica, se deberían implementar probablemente de un modo diferente. Otra importante cuestión, en el diseño de las tareas topológicas es cómo dos tareas topológicas pueden complementarse. Las limitaciones producidas por el sistema visual ortográfico son evitadas porque las tareas topológicas implementadas se complementan mutuamente.

Las tareas topológicas no tienen que ser definidas relativas a todo el cuerpo del robot o el agente, ya que hay la posibilidad de que se puedan definirse respecto a solo una región del cuerpo. A las distintas regiones sobre las que se puede definir tareas topológicas se les llamará *zonas*. En los agentes que se han implementado en los experimentos se han definido dos zonas. La zona 1 corresponde al cuerpo del agente y la zona 2 corresponde a la boca del agente. El uso de zonas es importante porque por ejemplo, la tarea *explorar* requiere establecer relaciones topológicas entre la zona 1 del agente y los objetos para que pueda sortearlos, mientras que la tarea *posicionarse para comer* requiere establecer relaciones topológicas entre la zona 2 del agente y los alimentos para que la boca quede enfrente de los alimentos.

##### 10.4.1.4.1 La tarea topológica de explorar

Entre todas las habilidades que un robot multifuncional necesita para ser efectivo, una de las habilidades más importantes es la capacidad de explorar. En este contexto,

“explorar” significa la habilidad de navegar en un entorno desconocido sin un destino específico. Esta habilidad es necesitada tanto por un robot moviéndose a través de las calles de una ciudad como por un robot que sea enviado a una misión de exploración de otro planeta. A continuación, se va a explicar cómo se le puede proporcionar a un robot la habilidad de explorar un entorno mediante una tarea topológica virtual. La implementación hecha en la investigación está destinada a espacios abiertos. El primer paso de la implementación de una tarea topológica es definir el número de metas generales de las que se compone la tarea. En este caso, la tarea topológica virtual consta de tres metas generales cuyos objetos son puertas. Como se ha mencionado, a las puertas se les asigna el valor  $-1$  para el campo *pattern*. Las puertas deben ser del mismo tamaño o más grandes que el agente lateralmente. Dado que el agente del micromundo tiene un ancho mayor que una unidad, solo tres de las seis posibles relaciones de tamaño vistas en la sección 10.2 pueden darse entre un agente y una puerta. La implementación que se ha programado es para un micromundo bidimensional, por lo tanto, las estructuras de datos contienen información sobre las dos posibles dimensiones. El campo de la estructura de datos *dimension 1* contiene información sobre la dimensión lateral del agente, y el campo *dimension 2* contiene información sobre la dimensión perpendicular al agente. Recuérdese que el sistema de referencia es egocéntrico y por esta razón las dimensiones son relativas al agente. Cada meta genérica determina la relación que debe ser establecida en cada dimensión. Podría darse el caso de que la tarea no establezca ninguna relación concreta para una dimensión. Eso significaría que son válidas en esa dimensión. Dependiendo de la distancia que se desee dejar entre los objetos reales y el agente al navegar, se determinarán unas relaciones topológicas u otras entre las puertas y el agente. Antes de implementar el código de una tarea topológica que usa puertas debe conocerse la estrategia para generar puertas que ha sido escogida en el PAE. En la tarea de exploración el PAE usa la estrategia de generar puertas en los límites de los sensores del agente. Esto lleva a que si el objetivo es atravesar una de las puertas el agente estará siempre navegando. Si hay obstáculos, el movimiento del agente para alinearse con las puertas provoca que a la vez evite los obstáculos. La figura 10.7 muestra los pasos de comportamiento mencionados. La implementación llevada a cabo de la tarea topológica puede ser consultada en el apéndice C. Así, la tarea topológica virtual que el agente almacena con el código del apéndice C causa que un agente explore constantemente el micromundo mientras evita obstáculos.

#### 10.4.1.4.2 La tarea topológica de *posicionarse para comer*

Se acaba de explicar cómo la tarea topológica de explorar hace al agente explorar el entorno, pero solo esa tarea no es suficiente para que el agente pueda sobrevivir. El agente necesita localizar alimento y posicionarse en una posición específica relativa al alimento para poder alimentarse. En este caso, la tarea topológica *posicionarse para comer* debe tener mayor prioridad que la tarea *explorar*. Por lo tanto, el agente deja

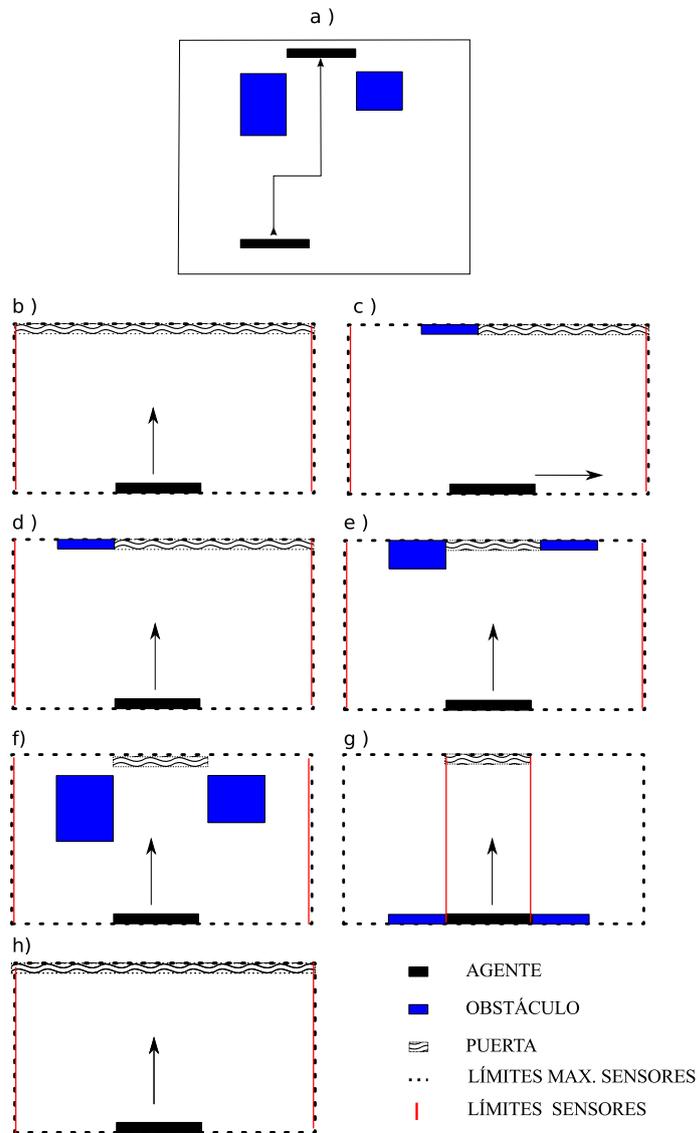


Figura 10.7: Ejemplo de cómo un agente con ANCT 1.0 lleva a cabo la tarea de explorar un entorno donde hay dos obstáculos. a) La trayectoria completa del agente en el entorno debido a la tarea topológica *explorar*. b) El agente avanza porque hay una puerta. c) Aparece un obstáculo lo que produce una puerta. El agente se mueve para estar enfrente de la puerta. d) Ahora el agente está enfrente de la puerta y avanza hacia ella. e) Aparecer otro obstáculo, pero la puerta es lo suficientemente ancha para que el agente continúe avanzando f) El agente continúa avanzando. g) El agente está a punto de dejar atrás los obstáculos. h) No hay obstáculos y la anchura de la puerta ocupa todo el horizonte, así el agente continúa avanzado.

de explorar cuando detecta alimento para ubicarse donde se encuentra el alimento. La implementación que se ha hecho hace que el agente use la tarea topológica *explorar* hasta que el agente detecta alimento. Cuando el agente detecta alimentos, al tener el alimento asociada una tarea topológica de mayor prioridad, abandona la tarea *explorar* y comienza la tarea *posicionarse para comer*. En la implementación a todos los objetos reales que son alimento se les asigna el valor 4 en el campo *pattern*. Así, la tarea topológica *posicionarse para comer* es iniciada si se registra un objeto con valor 4, y el agente maniobrará hasta que su boca esté enfrente del objeto. La tarea topológica *posicionarse para comer* contiene solo una meta general que contiene la información que ha sido mencionada. El apéndice C muestra el código que se ha implementado para especificar esta tarea topológica.

Sin embargo, el comportamiento que se acaba de describir puede no ser suficiente para que una gente tenga éxito sobreviviendo, ya que el agente podría quedar bloqueado indefinidamente por un obstáculo cuando maniobra para colocarse frente al alimento porque su comportamiento está enfocado completamente en ese objetivo y no contempla el tener que superar un obstáculo. Así, la implementación que se ha hecho, determina que cuando se encuentra un objeto con valor 4 en su campo “pattern” se comprueba si existe algún obstáculo que bloquee el camino al alimento; si no lo hay, entonces combinando la información de la meta general con el objetivo específico, se genera una meta específica. El código puede ser consultado en el apéndice C. En el caso de que se detecten obstáculos, la arquitectura usa de nuevo la técnica de crear puertas para evitar el obstáculo. Este método es usado recursivamente cada vez que se supera un obstáculo hacia el alimento. Así, en un momento determinado, el agente habrá evitado todos los obstáculos y no detectará más obstáculos entre él y el alimento y se generará la meta específica para situarse enfrente del alimento. La figura 10.8 muestra los pasos del comportamiento que se ha explicado.

#### 10.4.2. La ANCT 1.0, su implementación y prueba

Este apartado presenta los tipos de estructura de datos que usa la ANCT 1.0, y su implantación en agentes. Como ya se ha mencionado, la arquitectura ha sido diseñada con el objetivo de proporcionar a un agente la habilidad de navegar en un espacio abierto desconocido y dinámico usando la HSCT. En el primer y segundo apartado se mostrarán las estructuras de datos que se han usado para programar ANCT 1.0. El tercer apartado está dirigido a explicar el motor de inferencia. El cuarto apartado explica la necesidad del sistema de referencia egocéntrico para un agente que está en un entorno desconocido y dinámico. Finalmente, el último apartado de esta sección trata sobre la actuación de la ANCT 1.0 en un micromundo, ya que en la investigación se implementó la arquitectura en agentes que tienen que sobrevivir en un micromundo.

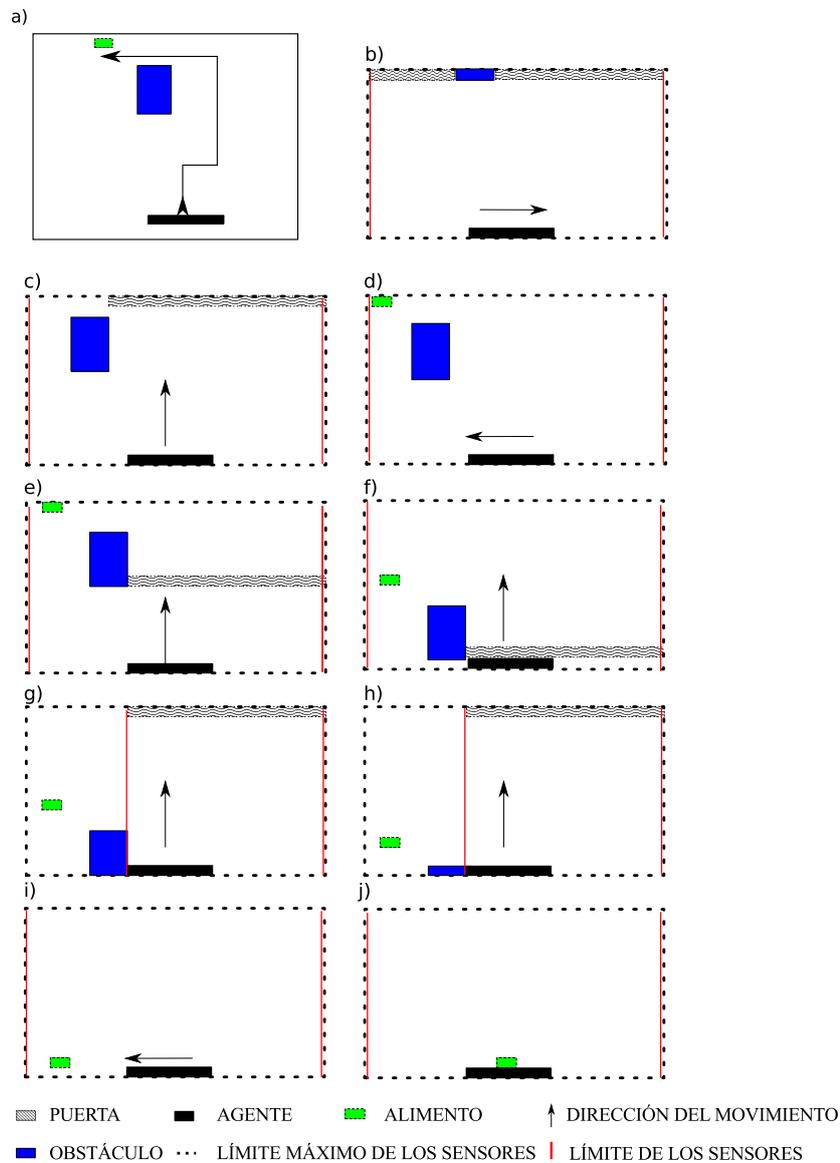


Figura 10.8: Ejemplo de cómo un agente con ANCT 1.0 ejecuta la tarea topológica *posicionarse para comer* habiendo un obstáculo para llegar al alimento. a) Muestra la trayectoria completa del agente en el entorno. b)c) El agente está ejecutando la tarea *exploración*. d) El agente encuentra un alimento, comprueba si hay o no un obstáculo, encuentra que si hay uno. e) El agente activa el modo esquivar, crea puertas, y se mueva a la puerta junto al obstáculo. f) El agente no ve el alimento y vuelve a la tarea *explorar*. g) El agente avanza. h) El agente vuelve a detectar el alimento pero ahora no detecta ningún obstáculo, así que comienza la tarea *posicionarse para comer*. i) El agente llega a la relación horizontal requerida. j) El agente llega a la relación vertical requerida.

### 10.4.2.1. Implementación de la representación de objetos

La implementación de la representación espacial que tiene la ANCT 1.0 está basada en la generalización de HSCT usada para un espacio n-dimensional. Así, los objetos en el espacio son descritos por un intervalo para cada dimensión espacial. Así, específicamente, el siguiente código de Object Pascal ha sido usado para implementar las representaciones de objetos en la ANCT 1.0:

```
Interval = Register
  first :integer;// begin of the interval
  last : integer;// end of the interval
end;

RObject = Register
  pattern: integer; // class of object assigned through the analysis
                    // of the information captured by the sensors
  distance: integer; // distance between the agent and the object
  dimensions: array [1..2] of Interval;
end;
```

En la estructura de datos, la diferencia entre objetos reales y virtuales viene dada por los valores en el campo *pattern*. Los valores negativos son reservados para clases de objetos virtuales y los números positivos son usados para clases de objetos reales. Los objetos virtuales son generados de la información recogida del entorno. En el caso de las puertas, la información es analizada de izquierda a derecha, localizando los espacios virtuales entre los objetos a la profundidad que la estrategia escogida necesita. A las puertas siempre se las asigna una unidad de profundidad, y el valor  $-1$  en el campo *pattern* para diferenciales de los objetos reales.

### 10.4.2.2. Las estructuras de datos de las tareas topológicas

En la implementación que se ha llevado a cabo de la ANCT 1.0 hay cuatro estructuras de datos principales para especificar tareas topológicas: TopologicalTask, TgeneralGoal, TgoalDimension y TconcreteGoal. El tipo de estructura de datos TopologicalTask se corresponde con el concepto de tarea topológica que se ha presentado anteriormente. Recuérdese que las metas generales son lo que indica a un agente lo que se debería de hacer con los objetos que pertenecen a una categoría particular. El tipo TopologicalTask permite el almacenamiento de variables del tipo TgeneralGoal. El tipo TgeneralGoal se corresponde con el concepto de meta general. A su vez, el tipo TgeneralGoal almacena variables del tipo TgoalDimension. El tipo TgoalDimension permite almacenar la relación espacial que se quiere establecer en una dimensión concreta. El tipo TconcreteGoal permite crear variables para almacenar la meta que debe llevarse a

cabo con un objeto específico. El tipo `TconcreteGoal` permite el almacenamiento de las relaciones que se deben establecer usando también la estructura `TgoalDimension`. Una cuestión importante que hay que tener en cuenta es que el cuerpo del agente puede ser dividido en zonas para poder especificar una tarea topológica respecto de una zona concreta del agente. El definir zonas en el cuerpo de un agente aporta una flexibilidad que a menudo es necesitada en la definición de tareas. El código de las estructuras de datos que se acaban de mencionar se puede encontrar en el apéndice C.

### 10.4.2.3. El motor de inferencia

Este apartado describe el motor de inferencia de la ANCT 1.0 en detalle. El motor de inferencia que ha sido desarrollado recibe la lista de objetos reales y virtuales generados por el PAE, y tiene acceso a la base de tareas topológicas. El motor de inferencia es un proceso que compara la lista de objetos reales con la base de tareas topológicas para ver si hay algún objeto real que se corresponda con una de las tareas topológicas reales. Si no se encuentra ningún objeto real que pueda ser emparejado con una tarea topológica real, la lista de puertas es confrontada con las tareas topológicas de exploración. Si hay puertas que encajen con alguna de las metas topológicas generales, entonces se seleccionará la puerta más cercana. En el caso de que se encontrara un objeto real con el cual una tarea topológica puede ser asociada, entonces comprueba si hay algún obstáculo entre el agente y el objeto. Como se ha mencionado antes, la localización de obstáculos es importante por la razón específica de que, aunque el agente plasme la información del entorno colocando objetos en una representación del espacio, la HSCT solo considera un objeto en su proceso de toma de decisiones en cada ejecución. Así, antes de preguntar a HSCT que acción llevar a cabo, es necesario considerar si la HSCT debe ser usada para tomar una decisión con el objeto o con un obstáculo. Para evitar obstáculos, la ANCT 1.0 usa una estrategia diferente si se trata de la tarea *explorar* o en la tarea *posicionarse para comer*. En la tarea *posicionarse para comer* se selecciona el obstáculos más cercano al agente; se generan las puertas a la altura del principio del obstáculo; se escoge la puerta pegando al obstáculo para generar la meta específica de atravesar la puerta fijando la relación espacial que está al otro lado. La realización de esa meta implica el evitar el obstáculo.

La localización de obstáculos, se lleva a cabo mediante dos procesos. El primer proceso encuentra un conjunto de objetos que son potencialmente obstáculos. Para encontrar ese grupo de objetos, se emplea la distancia entre el agente y el objeto objetivo. Todos los objetos reales que se encuentren a una distancia igual o menor que el objeto objetivo son obstáculos potenciales. Si el conjunto de obstáculos no es vacío, se lleva a cabo un segundo proceso que estudia la relación topológica del agente con cada uno de los obstáculos potenciales. Si encuentra un obstáculo potencial con el que la relación topológica no es disjuntos en la dimensión que existe lateralmente respecto del agente, a ese objeto se le considerará un obstáculo. Un ejemplo de cómo

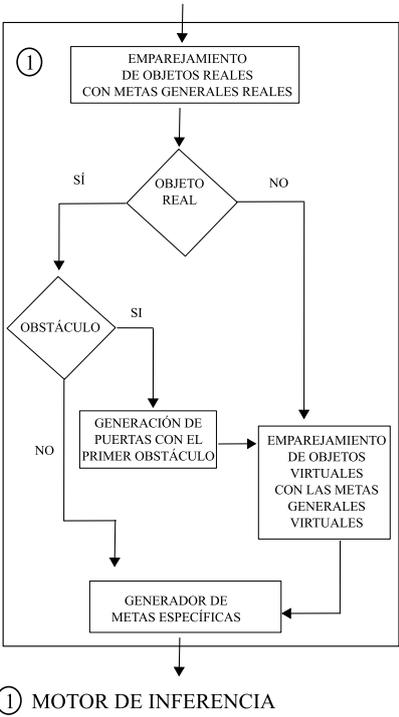


Figura 10.9: Diagrama del motor de inferencia.

actúa la ANCT 1.0 cuando encuentra obstáculos es dado en la figura 10.7. En caso de que no se detecten obstáculos, el motor de inferencia genera una meta específica con la información del objeto real. En la secuencia de decisiones que muestra la figura 10.8, se debe notar que la arquitectura no usa planificación sino que la acción que se lleva a cabo en cada momento es la reacción del agente a la situación en la que se encuentra en cada momento. Por ello, el agente es capaz de navegar en entornos desconocidos o dinámicos. Nótese que un entorno dinámico puede ser visto como un entorno desconocido en cada instante de tiempo. Por lo tanto, la ANCT 1.0 es capaz de operar en entornos dinámicos se están moviendo dado que cada acción que lleva a cabo un agente con la ANCT 1.0 es consecuencia solo de la información disponible en cada instante de tiempo.

#### 10.4.2.4. Percepción en ANCT 1.0

La HSCT se usó como un método de planificación en las anteriores investigaciones del programa RMMNT; pero ANCT 1.0 no tiene una representación del mundo manipulable, ya que no calcula ningún camino en una representación del mundo. La idea de arquitecturas sin representaciones del mundo fue propuesta por Rodney Brooks en las arquitecturas de subsunción [87]. La ANCT 1.0 tiene una percepción del mundo y reacciona a ella generando una acción como ya se ha mencionado. Tal vez, alguien podría hacerse las siguientes preguntas: ¿Qué ocurre con el conjunto de las representaciones de los objetos? ¿Es eso un mapa? La respuesta es que no lo es. Como mucho se puede hablar de un mapa local porque contiene información registrada por los sensores en tiempo real, y un mapa local no es útil porque solo permite planificar rutas muy cortas. Además, ANCT 1.0 no posee un mapa porque no tiene un proceso de planificación. Es decir, ANCT 1.0 no calcula posiciones futuras del agente en un mapa para crear rutas. Las estructuras de datos para representar los objetos en reactividad de segundo orden son el equivalente a las variables que almacenan si un parachoques ha sido o no activado en la reactividad de primer orden.

Otra cuestión a abordar es qué sistema de referencia debe usar una ANCT 1.0. La información espacial que un agente tiene, puede ser uno de estos dos tipos: alocéntrica o egocéntrica. En la información alocéntrica las coordenadas de los objetos son absolutas, y en la información egocéntrica las coordenadas son relativas a la posición del agente. En las anteriores investigaciones del programa RMMNT se ha supuesto que el agente era provisto de información alocéntrica teniendo un conocimiento total de las posiciones de los objetos y de él en el espacio. Sin embargo, la meta de la investigación presentada en esta sección ha sido enfrentarse al problema de no poder usar métodos de planificación al tener un entorno dinámico o desconocido, que son características de los entornos humanos. Así, el punto de partida de la investigación es que toda la información de la que se dispondría inicialmente es siempre relativa a la posición del agente. El hecho de que el agente tiene información alocéntrica o egocéntrica es importante

para implementar la HSCT porque las acciones son representadas de manera diferente dado que las acciones son representadas diferentemente, y, por lo tanto, tienen que ser etiquetadas de forma diferente.

Supóngase que hay un agente con información espacial egocéntrica en un espacio unidimensional con un objeto a su derecha. Si el agente tiene información espacial egocéntrica y se mueve a la derecha, el agente detectará que el objeto está más cerca de él que antes . Si el agente tiene información espacial alocéntrica, moviéndose a la derecha el agente detectará que él está más a la derecha que antes. Así, cuando la información es alocéntrica las acciones son representadas por funciones que modifican la posición del agente. Sin embargo, cuando la información es egocéntrica las acciones que el agente puede realizar son representadas como funciones que modifican la posición del objeto (u objetos) en el entorno que rodea al agente, pero no la posición del propio agente. Si hay un agente,  $A$ , que puede realizar tres acciones para moverse en un espacio unidimensional (avanzar, retroceder, permanece inmóvil), la representación de las acciones dependiendo de los sistemas de referencia se pueden ver en la tabla 10.5.

acción	función alocéntrica	definición	función egocéntrica	definición
1	$f_1$	$f_1 _x(x) = x + 1$	$\widehat{f}_1$	$\widehat{f}_1 _x(x) = x - 1$
2	$f_2$	$f_2 _x(x) = x$	$\widehat{f}_2$	$\widehat{f}_2 _x(x) = x$
3	$f_3$	$f_3 _x(x) = x - 1$	$\widehat{f}_3$	$\widehat{f}_3 _x(x) = x + 1$

Tabla 10.5: Representación de las acciones de  $o_A$

La función para etiquetar las acciones es la siguiente:

$$etiqueta(g) = \begin{cases} < & \text{if } \forall x \in X \quad x < g(x) \\ = & \text{if } \forall x \in X \quad x = g(x) \\ > & \text{if } \forall x \in X \quad x > g(x) \end{cases}$$

Tomando en consideración la tabla y la función anteriores, el etiquetado de las acciones de un agente  $A$  con la función de etiquetado genera la tabla 10.6.

función alocéntrica	etiqueta	función egocéntrica	etiqueta
$f_1$	<	$\widehat{f}_1$	>
$f_2$	=	$\widehat{f}_2$	=
$f_3$	>	$\widehat{f}_3$	<

Tabla 10.6: Tabla con las funciones etiquetadas.

Dado que las etiquetas de las acciones son diferentes si el agente tiene información alocéntrica o egocéntrica del espacio, un agente no puede tener información egocéntrica

y las acciones estar etiquetadas alocéntricamente porque eso conduciría a tomas de decisiones erróneas por la HSCT. Dado que el objetivo del programa RMMNT es contribuir a desarrollar sistemas capaces de funcionar autónomamente en entornos desconocidos, la mejor opción es usar información egocéntrica en la ANCT 1.0 ya que es la información que se obtiene de los sensores. Esta elección tiene las ventajas de que evita gastar tiempo transformando las coordenadas a alocéntricas y la acumulación de error en la posición del robot debida a problemas odométricos.

#### 10.4.2.5. Pruebas con la ANCT 1.0.

Con el objetivo de probar la ANCT 1.0, la implementación que se ha estado explicando de llevo a cabo para los agentes de un mundo virtual programado con DELPHI XE7. El micromundo es el mismo que se ha mencionado antes, pero se añadieron algunas características al micromundo. Una de esas nuevas características es que el agente necesita “comer” un tipo concreto de objeto del micromundo para permanecer vivo, y existen obstáculos de distintos tamaños. Los agentes tienen brazos que pueden mover para recoger cosas y una “boca” para comer cosas. Los obstáculos y los elementos comestibles son generados aleatoriamente al inicio de la ejecución del programa. La comida está solo disponible en su posición durante un corto periodo de tiempo, así que los agentes necesitan explorar el entorno constantemente y la probabilidad de que un comestible vuelva a surgir en la misma posición es muy pequeña. Los agentes no tienen un mapa con información sobre dónde están los comestibles y los obstáculos en el entorno. Los agentes solo tienen la información recogida por los sensores. El sistema de visión de los agentes les permite capturar los objetos, su anchura y la distancia de los objetos al agente. Esa información puede ser obtenida por un sistema comercial como Kinect.

De esta manera, el micromundo tiene muchas características de un entorno dinámico y desconocido. La ANCT 1.0 ha sido implantada en los agentes del micromundo, permitiéndoles obtener alimentos y sobrevivir en el mundo. El lector debería darse cuenta de que no solo la estructura es lo que hace a los agentes tener éxito en el entorno sino la combinación de la arquitectura con el conocimiento almacenado en ella. Ese conocimiento consiste en las dos tareas topológicas que han sido descritas anteriormente y que están contenidas en la base de tareas topológicas de la ANCT 1.0. Las simulaciones han mostrado que el agente puede sobrevivir indefinidamente en el entorno cuando los alimentos son suficientes en número y frecuencia.

## 10.5. Razonamiento cualitativo sobre orientación y giros

La sección anterior ha mostrado una arquitectura que permite la navegación en espacios desconocidos y dinámicos pero estos deben ser abiertos. El motivo es porque la versión inicial de HSCT fue concebida para un agente que se pudiera mover avanzando, retrocediendo, desplazarse a la izquierda y a la derecha. Ese conjunto de acciones (adelante, atrás, izquierda, y derecha) no permite establecer todas las posibles configuraciones para un agente porque ninguna permite cambiar la orientación ni ningún tipo de giro. Sin embargo, los desplazamientos hacia la izquierda y derecha pueden ser contruidos a partir de las acciones de avanzar, retroceder, y movimientos de giro. Por esta razón, muchos robots domésticos pueden moverse avanzando, retrocediendo y girando. Así, para superar la limitación de la navegación en espacios abiertos es necesario poder tomar decisiones sobre giros.

Girar y la dirección son cuestiones directamente relacionados. El razonamiento cualitativo sobre direcciones ha sido investigado en los campos de la cartografía y de los sistema de información geográfica(SSIG). El principal problema en la investigación en SSIG es cómo deducir la dirección de un punto A a un punto C, dada la dirección de A a B y de B a C[157]. Se han realizado diferentes propuestas sobre razonamiento tales como *cone-shaped directions* [157], *projection-based concept* [157], *double cross calculus* [161], *cardinal direction calculus* (CDC) [189], *flip-flop* [516], *dipole* [403], *OPRA*[402], *TPCC*[404], o *extended double cross*[356]. Esas propuestas abordan el problema de identificar qué objeto está siendo designado de un grupo de objetos, dado un conjunto de relaciones direccionales entre los objetos del grupo. Sin embargo, se ha hecho escasa investigación sobre navegación cualitativa mediante direcciones. Una propuesta para llevar a cabo razonamiento cualitativo sobre orientación está basada en la *Orient Point Relation Algebra* ( $OPRA_m$ ) con granularidad ajustable [136]. Otra propuesta ha sido el uso de un mapa de métrica-semántica-topológica [269]. Esta sección presenta los resultados que se han obtenido en el programa RMMNT sobre cómo nociones topológicas cualitativas pueden ser usadas para tomar decisiones de giro. El enfoque de la investigación que se presenta aquí difiere de los anteriores en que el proceso para tomar decisiones es una heurística, y la representación de la orientación se hace usando el conjunto de relaciones topológicas relativas.

El contenido de esta sección se estructura de la siguiente manera: el apartado 10.5.1 revisa los supuestos que se han asumido para realizar la investigación. El apartado explica el método hallado para codificar información direccional en una representación topológica 10.5.2. El apartado 10.5.3 presenta el método desarrollado para la toma de decisiones sobre giros, la heurística de semántica cualitativa direccional. El apartado 10.5.4 presenta los detalles de cómo se elige automáticamente el grafo de razonamiento topológico direccional. El apartado 10.5.5 presenta las pruebas que se han hecho con el

método desarrollado. Finalmente, el apartado 10.5.6 propone un sistema de coordenadas alternativo al que se usó en la sección anterior, y que está basado en los resultados obtenidos sobre giros.

### 10.5.1. Premisas iniciales

La investigación que se ha llevado a cabo ha sido considerando como espacio de navegación un espacio bidimensional euclídeo, donde hay dos posibles acciones: girar a la izquierda y a la derecha. Adicionalmente, se han hecho algunas premisas adicionales relativas a los objetos y el agente, para enfocar la investigación en situaciones concretas. Así, durante la investigación, cada situación considerada cumple las siguientes condiciones:

- El espacio bidimensional tiene una geometría euclídea.
- El agente es simétrico, y su punto de rotación es el centro de masas considerando que el agente tiene su masa distribuida uniformemente.
- Si el objeto es cóncavo no está rodeando el agente ni parcialmente ni completamente. Las cavidades de los objetos cóncavos deben ser menores que el agente, de manera que el agente no puede estar dentro de las cavidades del objeto respecto al cual se busca establecer una orientación. Esta condición es porque un agente en una cavidad está rodeado por el objeto, por lo que no se puede hablar sobre una relación direccional entre ambos. La situación de un agente rodeado por un objeto debería ser representada por una relación topológica como se discutió en la sección 10.3.
- El agente no puede colisionar con el objeto mientras está girando. Esto implica que la distancia más corta del objeto al agente es mayor que la distancia entre el centro de rotación del agente y el punto más alejado del agente al centro de rotación.

La última premisa del conjunto es porque para enfocar la investigación en describir relaciones de orientación relativa se evita la posibilidad de que haya solapamiento entre el agente y el objeto mientras el agente gira. Así, la cuestión de detectar colisiones durante el giro queda fuera de los objetivos de la investigación.

### 10.5.2. Transformando información direccional en información topológica

El primer elemento de una heurística es una representación de los posibles estados cualitativos del problema. La HSCT usa relaciones cualitativas para representar los

estados, que pueden ser entendidas como coordenadas cualitativas cartesianas porque el conjunto de relaciones que se usan para espacios de dimensiones mayores a uno se crea mediante el producto cartesiano del conjunto de relaciones para una dimensión. Sin embargo, esas relaciones no permiten razonar sobre giros. Así, si se desea girar se necesita un sistema de coordenadas cualitativo diferente para tomar decisiones. El punto de partida en la investigación son los dos siguientes hechos:

- Un agente con una simetría continua no modifica su posición espacial cuando gira sobre sí mismo.
- Las acciones de giro modifican la dirección entre un agente y un objeto, incluso si el agente tiene una simetría continua.

En base a los dos hechos anteriores, se dedujo que no se podía usar solo información topológica para hacer decisiones sobre girar, y que la información direccional es esencial para razonar sobre giros. Por lo tanto, el camino que se ha explorado ha sido codificar información direccional en una representación topológica para luego tomar decisiones en la representación topológica. Dado que la dirección relativa entre un agente y un objeto es alterada cuando un agente gira, la meta ha sido desarrollar un método que exprese cómo y en qué dirección el objeto y el agente están alineados. El método encontrado para representar el alineamiento consiste en proyectar el objeto y el agente en un espacio unidimensional que es paralelo al eje  $x$  del sistema de coordenadas egocéntrico fijado en el centro de rotación del agente. La orientación del sistema de coordenadas usa como eje  $y$  la dirección de avance del agente. Una vez las proyecciones son obtenidas, se calcula la relación topológica relativa entre el agente y el objeto. Este proceso es plasmado en la figura 10.10. El agente de la figura 10.10 no posee una simetría continua porque una forma rectangular permite mostrar mejor los cambios de los giros, pero el método funciona también con objetos con simetría continua.

Los sensores de un robot registran información cuantitativa, y el agente puede representar esta información usando un sistema egocéntrico de coordenadas. El eje vertical del sistema egocéntrico de coordenadas corresponde con la dirección en la cual el agente puede moverse avanzando y retrocediendo. El origen de coordenadas es el punto de rotación del agente, y los ejes verticales y horizontales son ortogonales. Por lo tanto, el método usado para calcular la proyección del sistema está tomando las coordenadas  $x$  del objeto y del agente en el sistema de coordenadas egocéntrico. El sistema de coordenadas debe ser egocéntrico porque esto permite conocimiento de la alineación causada por el movimiento de giro del agente. Si el sistema de referencia fuera aloécéntrico, las coordenadas y los ejes de coordenadas serían estáticos, y el objeto mantendría los mismos valores independientemente de los giros que haga el agente. Por lo tanto, no se registraría ningún cambio porque el origen de la proyección sería siempre el mismo. La figura 10.11 muestra los casos de giro de un agente con un sistema de coordenadas egocéntrico y aloécéntrico.

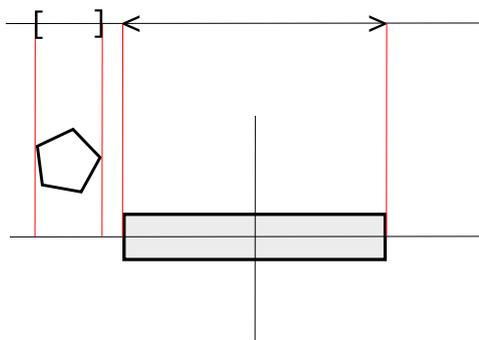


Figura 10.10: Proyección de un agente y un objeto en un espacio unidimensional. El pentágono es proyectado a un espacio unidimensional, generando  $[ \ ]$  y la proyección del agente genera el objeto  $< \ >$ . La proyección es paralela al eje horizontal del sistema egocéntrico de coordenadas del agente.

La proyección permite la identificación de una relación topológica que codifica información sobre la dirección relativa entre el objeto y el agente. La proyección permite la identificación de una relación topológica que codifica información direccional relativa a el objeto y el agente. Sin embargo, debido a que la proyección es de un espacio bidimensional a un espacio unidimensional, hay alguna información que se pierde. Esa pérdida de información puede verse en la figura 10.12. La proyección da información sobre el alineamiento, pero no es posible saber en qué dirección el alineamiento ocurre.

Aunque en la investigación no se consideró otras acciones aparte del giro, la pérdida de información es catastrófica para la elección entre acciones de avanzar y retroceder, por lo que debe de encontrarse una manera de que no se pierda. Para lograr esa meta, se ha usado la idea de crear variantes posicionales de las relaciones topológicas relativas. Concretamente se han definido tres variantes posicionales de cada relación topológica relativa. Cada variante posicional contiene información sobre la dirección donde la proyección fue hecha. Las variantes posicionales son denotadas por los símbolos:  $+$ ,  $\approx$ ,  $-$ . Las variantes posicionales forman un conjunto, que es denotado por  $PV$ . Usando los conjuntos  $S_{13}$  y  $PV$ , se define un nuevo conjunto de relaciones de la siguiente manera:

$$D_{39} = S_{13} \times PV \text{ donde } PV = \{+, \approx, -\}$$

A las relaciones del conjunto  $D_{39}$  se las denomina *relaciones topológicas direccionales* (RRTTDD). Las variantes posicionales vienen de la división del espacio bidimensional en tres regiones desde las que la proyección de los objetos se lleva a cabo. Esas tres regiones son relativas al agente porque el sistema de coordenadas es egocéntrico. El espacio bidimensional es dividido, mediante dos líneas rectas paralelas al eje  $x$ : Una corta al punto con mayor coordenada  $y$  del agente y la otra corta el punto del agente con menor coordenada  $y$ . Las regiones que surgen al trazar las dos líneas son mostradas

COORDINATE SYSTEM	1	2	3	4
EGOCENTRIC				
ALLOCENTRIC				

Figura 10.11: Un agente con forma rectangular gira con diferente tipo de coordenadas. Un agente está girando de 1 a 4. En el caso con coordenadas alocéntrico, la relación topológica de las proyecciones del objeto y el agente no cambiaría.

en la figura 10.13.

Para simplificar la notación de las relaciones topológicas direccionales, el par que denota una relación es sustituido por la relación topológica relativa del par con un superíndice que será el símbolo de la variante posicional del par. Por ejemplo:

$$(s_8, +) = s_8^+$$

La asignación de la variante posicional a la relación viene determinada por la región desde la que el objeto es proyectado. La regla para asignar una variante posicional es que si el objeto está completamente dentro de las zonas + o - esa será la variante posicional que se le asigne a la proyección; pero si alguna de las partes del objeto está dentro de la zona  $\approx$  entonces la variante posicional que se le asignará a la relación es  $\approx$ . Este método soluciona el problema de la pérdida de información como se muestra en la figura 10.14

### 10.5.3. Heurística de semántica cualitativa direccional

Como se ha mencionado en apartados anteriores, se está considerando agentes que tienen dos direcciones en las que girar, ya que se ha supuesto que están en un espacio bidimensional. Lo cierto es que sería suficiente un agente con una única acción para girar hasta establecer cualquier relación topológica direccional que pueda darse con un objeto porque las relaciones se suceden cíclicamente. Por lo tanto, establecer una relación topológica direccional con un objeto puede siempre ser resuelto usando la misma acción de giro. Esto ha llevado a proponer una meta para la heurística que se ha desarrollado, la de que la heurística no solo establezca una relación con un objeto, sino que elija la acción de giro que use menos tiempo y energía para establecerla.

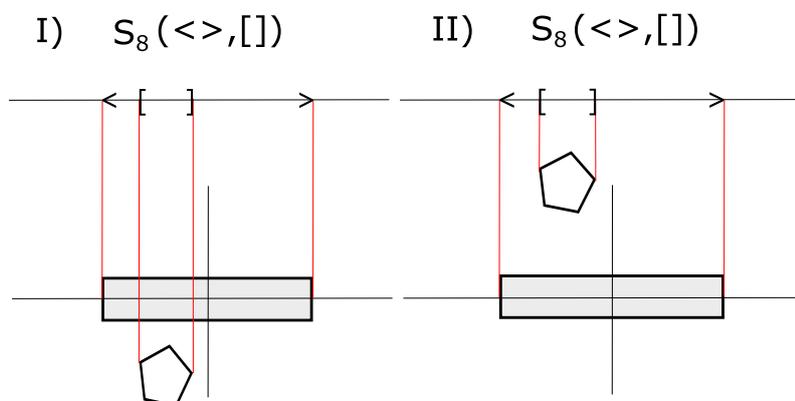


Figura 10.12: Un agente con una forma rectangular y un objeto con forma pentagonal son proyectados en un espacio unidimensional en dos casos con posiciones distintas. La situación I) no puede ser distinguida de la II) usando RTRs porque ambas situaciones generan la misma proyección. Este ejemplo muestra que hay una pérdida de información en la proyección del objeto y el agente en el espacio unidimensional.

Para lograr esa meta, es necesario una estructura que organice las representaciones y permita determinar los efectos de las decisiones. Las estructuras de orden total permiten el establecimiento de una heurística para sus acciones. A diferencia de las relaciones topológicas relativas, las relaciones topológicas direccionales no pueden ser organizadas como una estructura de orden total porque después de que un agente gira 360 grados, se encontrará de nuevo en la misma relación con el objeto. Así, la estructura elegida para organizar las relaciones topológicas direccionales es un grafo cíclico. Un grafo cíclico posee un orden cíclico. Un orden cíclico es una relación ternaria que satisface ciclicidad, asimetría, transitividad, y totalidad. Esas propiedades implican que el orden cíclico organice un conjunto de elementos en un círculo. Los nodos de un grafo cíclico son los elementos de  $D_{39}$ . Cada uno de esos grafos cíclicos es denominado *grafo de razonamiento topológico direccional* (GRTD). Las aristas de un GRTD conducen a la RTD anterior o siguiente entre un agente y un objeto cuando el objeto está girando. La dirección de las aristas encaja con la dirección del giro. Para representar un orden cíclico y los elementos del orden cíclico, se va a usar unos paréntesis, pero con las posiciones cambiadas al uso normal (imitando que son dos lados de un eslabón) para denotar que el último elemento está relacionado con el primero. Así, si  $C$  es el siguiente orden cíclico:

$$C = \{(e_1, e_2), (e_2, e_3), (e_3, e_1)\}$$

se denotará como:

$$)e_1, e_2, e_3($$

Como en los GLLRT, los GGRDT que pueden o no existir están determinados por los tamaños del agente y del objeto. El ángulo de giro es un parámetro cuyos

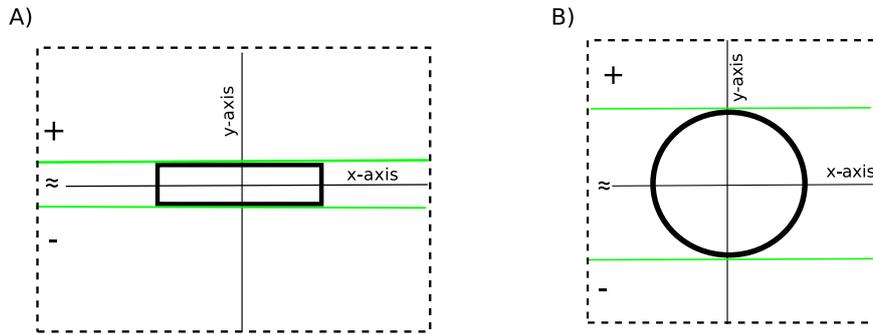


Figura 10.13: Con el objetivo de evitar la pérdida de información en las proyecciones el espacio bidimensional es dividido en tres regiones. Las líneas verdes denotan los límites entre las tres regiones. A) El agente tiene forma rectangular. B) El agente tiene forma rectangular.

dominios son los reales y por lo tanto continuo. Esto difiere de las situaciones estudiadas anteriores donde el espacio era discreto. Debido a que el espacio era discreto aparecían 6 situaciones diferentes con un GLRT para cada una. La continuidad del dominio del ángulo reduce el número de casos relativos al tamaño a 3. Sin embargo, la posición en el espacio del objeto respecto al agente causa la asignación de diferentes variantes posicionales. Las condiciones establecidas en el apartado 10.5.1 también determinan casos de secuencias de RRTD que pueden y no pueden ocurrir. Por ejemplo, al asumir una geometría euclídea, existe un GRTD en el que existe un nodo para  $s_6^-$ , pero no hay ninguno que tenga un nodo para  $s_7^-$  o  $s_8^-$ . Es decir, existen GRTD que no pueden existir debido a la condición de que el agente no puede colisionar con el objeto mientras está girando. Un ejemplo de GRTD que no puede existir, si el agente es más grande que el objeto, es el siguiente:

$$)s_1^-, s_2^-, s_3^-, s_7^+, s_{11}^-, s_{12}^-, s_{13}^-, s_{12}^-, s_{11}^-, s_7^-, s_3^-, s_2^-, ($$

La razón de que ese GRTD no pueda existir es porque la RTD entre un objeto y un agente no puede cambiar en un espacio euclídeo de  $s_3^-$  a  $s_7^+$  sin que ocurra una colisión entre el objeto y el agente al girar el agente.

Otro ejemplo de GRTD que no puede existir siendo el objeto más grande que el agente es el siguiente:

$$)s_1^-, s_2^-, s_3^-, s_4^-, s_6^-, s_9^-, s_{11}^-, s_{12}^-, s_{13}^-, s_{12}^-, s_{11}^-, s_9^-, s_6^-, s_4^-, s_3^-, s_3^-, s_2^-($$

La razón es porque si el objeto no está rodeando al agente, la RTD no puede cambiar de  $s_4^-$  a  $s_6^-$  directamente ni de  $s_6^-$  a  $s_9^-$ .

Existen múltiples GRTD que no son posibles, pero hay 15 que son posibles cuando se cumplen las condiciones que se han fijado en el apartado 10.5.1. Inicialmente, hay tres

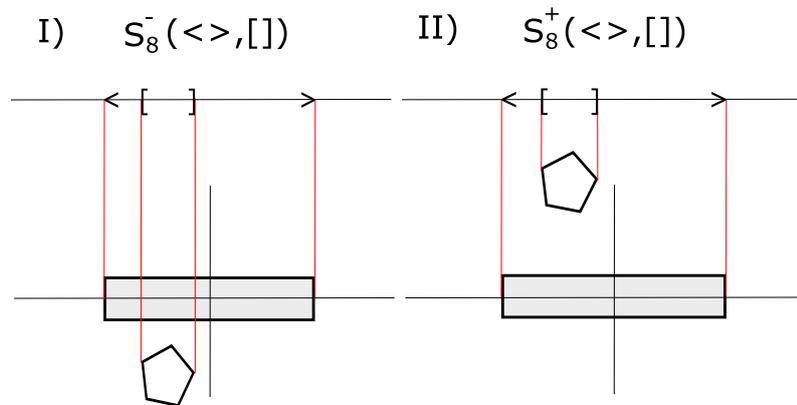


Figura 10.14: Un agente con una forma rectangular y un objeto conforma pentagonal son proyectados en el espacio unidimensional en dos situaciones diferentes. A diferencia de la asignación de la relación topológica relativa mostrada en la figura 10.12, asignando una relación topológica direccional se distinguen los casos I) y II).

casos principales a considerar: la proyección del objeto en el eje horizontal del agente es mayor que el agente, igual al agente y más pequeño que el agente. Los grafos que pertenecen a cada uno de esos casos se denotan como:  $GRTD^>$ ,  $GRTD^=$  y  $GRTD^<$ . Además, cada uno de los anteriores casos tiene variantes debido a la posición del objeto, lo que determina las quince variantes. Cada una de las variantes será indicada añadiendo un subíndice a los casos principales. Los quince GGRTD son listados en la tabla 10.7, y la figura 10.15 muestra la representación gráfica de un GRDT, específicamente el  $GRDT_1^=$ .

Id.	Grafo de razonamiento topológico direccional
$GRTD_1^>$	$)s_1^-, s_1^+, s_2^+, s_3^+, s_5^+, s_8^+, s_{10}^+, s_{11}^+, s_{12}^+, s_{13}^+, s_{13}^-, s_{13}^-, s_{12}^-, s_{11}^-, s_{10}^-, s_8^-, s_5^-, s_3^-, s_2^-, s_1^- ($
$GRTD_2^>$	$)s_1^-, s_2^+, s_3^+, s_5^+, s_8^+, s_{10}^+, s_{11}^+, s_{12}^+, s_{13}^-, s_{12}^-, s_{11}^-, s_{10}^-, s_8^-, s_5^-, s_3^-, s_2^- ($
$GRTD_3^>$	$)s_1^-, s_2^-, s_3^+, s_5^+, s_8^+, s_{10}^+, s_{11}^+, s_{12}^-, s_{13}^-, s_{12}^-, s_{11}^-, s_{10}^-, s_8^-, s_5^-, s_3^-, s_2^- ($
$GRTD_4^>$	$)s_1^-, s_2^-, s_3^-, s_3^+, s_5^+, s_8^+, s_{10}^+, s_{11}^+, s_{11}^-, s_{12}^-, s_{13}^-, s_{12}^-, s_{11}^-, s_{11}^-, s_{10}^-, s_8^-, s_5^-, s_3^-, s_3^-, s_2^- ($
$GRTD_1^<$	$)s_1^-, s_1^+, s_2^+, s_3^+, s_4^+, s_6^+, s_9^+, s_{11}^+, s_{12}^+, s_{13}^+, s_{13}^-, s_{13}^-, s_{12}^-, s_{11}^-, s_9^-, s_6^-, s_4^-, s_3^-, s_2^-, s_1^- ($
$GRTD_2^<$	$)s_1^-, s_2^+, s_3^+, s_4^+, s_6^+, s_9^+, s_{11}^+, s_{12}^+, s_{13}^-, s_{12}^-, s_{11}^-, s_9^-, s_6^-, s_4^-, s_3^-, s_2^- ($
$GRTD_3^<$	$)s_1^-, s_2^-, s_3^+, s_4^+, s_6^+, s_9^+, s_{11}^+, s_{11}^-, s_{12}^-, s_{13}^-, s_{12}^-, s_{11}^-, s_9^-, s_6^-, s_4^-, s_3^-, s_2^- ($
$GRTD_4^<$	$)s_1^-, s_2^-, s_3^-, s_3^+, s_4^+, s_6^+, s_9^+, s_{11}^+, s_{11}^-, s_{12}^-, s_{13}^-, s_{12}^-, s_{11}^-, s_{11}^-, s_9^-, s_6^-, s_4^-, s_3^-, s_3^-, s_2^- ($
$GRTD_5^<$	$)s_1^-, s_2^-, s_3^-, s_4^+, s_6^+, s_9^+, s_{11}^+, s_{12}^-, s_{13}^-, s_{12}^-, s_{11}^-, s_9^-, s_6^-, s_4^-, s_3^-, s_2^- ($
$GRTD_6^<$	$)s_1^-, s_2^-, s_3^-, s_4^+, s_6^+, s_9^-, s_{11}^-, s_{12}^-, s_{13}^-, s_{12}^-, s_{11}^-, s_9^-, s_6^-, s_4^-, s_3^-, s_3^-, s_2^- ($
$GRTD_7^<$	$)s_1^-, s_2^-, s_3^-, s_4^-, s_6^-, s_6^+, s_9^-, s_{11}^-, s_{12}^-, s_{13}^-, s_{12}^-, s_{11}^-, s_9^-, s_6^-, s_6^-, s_6^-, s_4^-, s_3^-, s_3^-, s_2^- ($
$GRTD_1^=$	$)s_1^-, s_1^+, s_2^+, s_3^+, s_7^+, s_{11}^+, s_{12}^+, s_{13}^+, s_{13}^-, s_{13}^-, s_{12}^-, s_{11}^-, s_7^-, s_3^-, s_2^-, s_1^- ($
$GRTD_2^=$	$)s_1^-, s_2^+, s_3^+, s_7^+, s_{11}^+, s_{12}^+, s_{13}^-, s_{12}^-, s_{11}^-, s_7^-, s_3^-, s_2^-, ($
$GRTD_3^=$	$)s_1^-, s_2^-, s_3^+, s_7^+, s_{11}^+, s_{12}^-, s_{13}^-, s_{12}^-, s_{11}^-, s_7^-, s_3^-, s_2^-, ($
$GRTD_4^=$	$)s_1^-, s_2^-, s_3^-, s_3^+, s_7^+, s_{11}^+, s_{11}^-, s_{12}^-, s_{13}^-, s_{12}^-, s_{11}^-, s_{11}^-, s_7^-, s_3^-, s_3^-, s_2^-, ($

Tabla 10.7: Esta tabla muestra los GGRTD que pueden existir entre un objeto y un agente que cumple las condiciones citadas en el apartado 10.5.1.

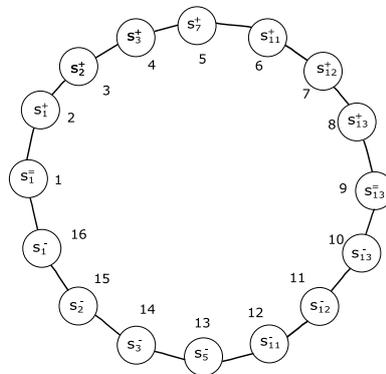


Figura 10.15: El grafo de la figura es el  $GRDT_1^=$  para un objeto y agente que tienen el mismo tamaño. Los nodos son etiquetados con números naturales de acuerdo a la función  $\mu$ .

Para determinar el camino más corto, el agente debe identificar el GRTD para tomar decisiones sobre giros. La identificación del correcto GRTD requiere conocer el tamaño del agente y el objeto; pero el tamaño máximo en el espacio bidimensional no es el valor que se necesita (como sí es en la HSCT) porque el tamaño máximo de la

proyección depende de la orientación entre el agente y el objeto. Se dedicará el apartado 10.5.4 específicamente a explicar el método para elegir el correcto GRTD para tomar decisiones.

Considérese de momento que se conoce el correcto GRTD para tomar decisiones. En esta situación, supóngase que tenemos un agente robótico con dos acciones de giro:  $a_1$  y  $a_2$ . La acción  $a_1$  hace girar a la izquierda al agente y la acción  $a_2$  le hace girar a la derecha. La heurística debe elegir que acción lleva a establecer la orientación que quiere establecer en menos tiempo. Para conseguirlo, el método que se ha desarrollado es el siguiente. Cada nodo de GRTD es etiquetado con un número natural. El etiquetado se realiza asignando 1 a cualquiera de los nodos del GRTD y se viaja a través del grafo siempre en la misma dirección asignando un nuevo número a cada uno de los nodos que es una unidad mayor al número asignado en el anterior nodo. El etiquetado del grafo se detiene cuando se llega a un nodo que ya está etiquetado. El etiquetado del grafo será representado con la función  $\mu$ . La figura 10.15 muestra el etiquetado de un GRTD. Además, cada una de las acciones del agente es etiquetada con la dirección que produce en el GRTD el cambio de la RTD cuando la acción es aplicada. Así,  $a_1$  es etiquetada con  $\curvearrowright$  y  $a_2$  con  $\curvearrowleft$ . El siguiente paso consiste en calcular la distancia en ambas direcciones desde el nodo que se corresponde con la RTD actual y el nodo correspondiente a la RTD deseada, y así, seleccionar la dirección que conlleva la menor distancia entre ambos nodos. Las distancias se calculan de la siguiente manera:

$$d_{\curvearrowright} = \begin{cases} \mu(n_w) - \mu(n_c) & \text{if } \mu(n_w) \geq \mu(n_c) \\ |N| - |\mu(n_w) - \mu(n_c)| & \text{if } \mu(n_w) < \mu(n_c) \end{cases}$$

$$d_{\curvearrowleft} = \begin{cases} |N| - |\mu(n_w) - \mu(n_c)| & \text{if } \mu(n_w) \geq \mu(n_c) \\ \mu(n_c) - \mu(n_w) & \text{if } \mu(n_w) < \mu(n_c) \end{cases}$$

donde  $n_c$  es el nodo correspondiente a la RTD actual,  $n_w$  es el nodo correspondiente a la RTD que se desea establecer,  $|N|$  es la cardinalidad del conjunto de nodos del GRTD, y  $|\mu(n_w) - \mu(n_c)|$  el valor absoluto de la resta indicada.

Una vez las distancias han sido calculadas, se usa la tabla 10.8 para recuperar una etiqueta de dirección, y la etiqueta que se recupera se usa a su vez para recuperar de la tabla 10.9 la acción de giro que el agente tomará.

Situation	$d_{\curvearrowright} > d_{\curvearrowleft}$	$d_{\curvearrowright} < d_{\curvearrowleft}$	$d_{\curvearrowright} = d_{\curvearrowleft} \neq 0$
Label	$\curvearrowright$	$\curvearrowleft$	any of both

Tabla 10.8: Tabla de distancias y direcciones.

Action	Label
$a_1$	$\curvearrowright$
$a_2$	$\curvearrowleft$

Tabla 10.9: Tabla de direcciones y acciones.

Se debe notar que si  $d_1 = d_2 = 0$ , la RTD que se buscaba establecer ha sido establecida.

#### 10.5.4. La elección del grafo de razonamiento topológico direccional

En el apartado anterior se ha mencionado brevemente el paso de elegir el GRTD correcto para la toma de decisiones. La complejidad de elegir el correcto GRTD emerge de la forma de los objetos, ya que es posible que una parte del objeto quede oculta al sistema visual del agente en algunas posiciones, y el agente no pueda saber la longitud máxima de la proyección del objeto si solo tiene la información que le aporta su sistema visual. Si el agente no conoce el tamaño máximo de la proyección del objeto, no puede saber que RTD ocurrirá cuando el agente gire. Elegir el GRTD correcto para tomar decisiones es un callejón sin salida porque no se tiene suficiente información visual. Así, el problema a abordar es cómo el agente puede adquirir suficiente información para elegir el GRTD correcto.

La solución que se propone desde la investigación realizada requiere una estructura física específica en el agente. El agente debe poseer un mecanismo de “cuello” que permita girar su sistema visual mientras el resto del agente permanece en su posición. Girar el sistema visual gasta menos energía que girar el agente completo y evita problemas innecesarios de estabilidad. Al girar su sistema visual, el agente obtiene información sobre el GRTD correcto de la siguiente manera. Mientras el agente está girando su sistema visual, registra a la vez que RRTTDD se están estableciendo. Entre las RRTTDD, hay algunas que solo existen en uno de los GRTD. Así, el agente gira su sistema visual hasta encontrar una RTD que solo está en un GRTD. El sistema para denominar los GGRTD establece tres casos principales y variantes de cada uno de los casos principales. El superíndice  $\square$  determina el caso principal, y el subíndice determina la variante del caso. Así, un GRTD desconocido se representa como  $GRTD_i^\square$ . Un GRTD tiene dos componentes: el primero una relación topológica relativa (consúltese 10.2) y el segundo una variante posicional. El caso principal puede ser determinado detectando una RTR que diferencie un caso de otro. A cada una de esas RRTTRR se la denominará *clave*. Las RRTTRR clave que determinan el caso principal del DTRG correcto para la toma de decisiones son:

- *cubre* – 0,  *contiene* o  *cubre* – 1 determinan  $DTRG^>$

- *iguales* determina  $DTRG^=$
- *cubierto - 0*, *contenido* o *cubierto - 1* determinan  $DTRG^<$

Una vez se ha determinado el caso principal, se puede determinar la variante detectando dos RRTTDD. Cada uno de los pares de RRTTDD que determina una variante será denominado *par clave*. Por lo tanto, el algoritmo para determinar un  $GRTD_i^\square$  para tomar decisiones tiene dos pasos principales. Primero, el agente gira su sistema visual hasta que encuentre una RTR clave con el objeto. Esa RTR clave determinará el valor de  $\square$  del  $GRTD_i^\square$ . La dirección de giro del sistema visual necesitada para encontrar una RTR clave depende de la RTD inicial entre el objeto y el agente. Si el sistema visual tiene una amplitud de 360 grados de visión y la RTD inicial no es una clave, el algoritmo para girar el sistema visual consta de las siguientes cuatro reglas:

- Si la variante posicional es  $+$  o  $\approx$ , y la relación topológica relativa es *disjuntos-0*, *pegados - 0* o *solapados - 0*, el sistema visual gira a la izquierda.
- Si la variante posicional es  $+$  o  $\approx$ , y la relación topológica relativa es *disjuntos-1*, *pegados - 1* o *solapados - 1*, el sistema visual gira a la derecha.
- Si la variante posicional es  $-$  y la relación topológica relativa es *disjuntos - 0*, *pegados - 0* o *solapados - 0*, el sistema visual gira a la derecha.
- Si la variante posicional es  $+$  y la relación topológica relativa es *disjuntos - 1*, *pegados - 1* o *solapados - 1*, el sistema visual gira a la izquierda.

Si el sistema visual solo tiene una amplitud de 180 grados de visión, el camino para encontrar la RTR clave será más largo que en los casos donde la variante posicional inicial de la relación entre el objeto y el agente sea  $-$ , y las reglas que se seguirán serán las siguientes:

- Si la variante posicional es  $+$ ,  $\approx$ , o  $-$ , y la relación topológica relativa es *disjuntos-0*, *pegados - 0* o *solapados - 0* el sistema visual gira a la izquierda.
- Si la variante posicional es  $+$ ,  $\approx$ , or  $-$ , y la relación topológica relativa es *disjuntos-1*, *pegados - 1* o *solapados - 1* el sistema visual gira a la derecha.

El segundo paso es detectar un par clave. La detección del par clave puede ser realizada cuando el agente gira el sistema visual en la dirección inversa a la que lo ha hecho en el primer paso, para devolverlo a su posición inicial. Si la RTD es una clave, entonces el sistema visual debe girar específicamente para detectar el par clave. El sistema visual necesita girar un máximo de 90 grados para detectar un par clave. La tabla 10.10 muestra cuáles son los pares claves que permiten determinar cada una de las variante cuando se conoce el caso principal.

variante	Par clave
$DTRG_1^>$	$(s_1^-, s_1^+), (s_{13}^+, s_{13}^-), (s_{13}^-, s_{13}^-), (s_1^-, s_1^-)$
$DTRG_2^>$	$(s_1^-, s_2^+), (s_{12}^+, s_{13}^-), (s_{13}^-, s_{12}^-), (s_2^-, s_1^-)$
$DTRG_3^>$	$(s_2^-, s_3^+), (s_{11}^+, s_{12}^-), (s_{12}^-, s_{11}^-), (s_3^-, s_2^-)$
$DTRG_4^>$	$(s_3^-, s_3^+), (s_{11}^+, s_{11}^-), (s_{11}^-, s_{11}^-), (s_3^-, s_3^-)$
$DTRG_1^<$	$(s_1^-, s_1^+), (s_{13}^+, s_{13}^-), (s_{13}^-, s_{13}^-), (s_1^-, s_1^-)$
$DTRG_2^<$	$(s_1^-, s_2^+), (s_{12}^+, s_{13}^-), (s_{13}^-, s_{12}^-), (s_2^-, s_1^-)$
$DTRG_3^<$	$(s_2^-, s_3^+), (s_{11}^+, s_{12}^-), (s_{12}^-, s_{11}^-), (s_3^-, s_2^-)$
$DTRG_4^<$	$(s_3^-, s_3^+), (s_{11}^+, s_{11}^-), (s_{11}^-, s_{11}^-), (s_3^-, s_3^-)$
$DTRG_5^<$	$(s_3^-, s_4^+), (s_9^+, s_{11}^-), (s_{11}^-, s_9^-), (s_4^-, s_3^-)$
$DTRG_6^<$	$(s_4^-, s_6^+), (s_6^+, s_9^-), (s_9^-, s_6^-), (s_6^-, s_4^-)$
$DTRG_7^<$	$(s_6^-, s_6^+), (s_6^+, s_6^-), (s_6^-, s_6^-), (s_6^-, s_6^-)$
$DTRG_1^-$	$(s_1^-, s_1^+), (s_{13}^+, s_{13}^-), (s_{13}^-, s_{13}^-), (s_1^-, s_1^-)$
$DTRG_2^-$	$(s_1^-, s_2^+), (s_{12}^+, s_{13}^-), (s_{13}^-, s_{12}^-), (s_2^-, s_1^-)$
$DTRG_3^-$	$(s_2^-, s_3^+), (s_{11}^+, s_{12}^-), (s_{12}^-, s_{11}^-), (s_3^-, s_2^-)$
$DTRG_4^-$	$(s_3^-, s_3^+), (s_{11}^+, s_{11}^-), (s_{11}^-, s_{11}^-), (s_3^-, s_3^-)$

Tabla 10.10: Pares clave para detectar las variantes de GRTD cuando el agente gira a la derecha su sistema visual. Si el agente gira a la izquierda su sistema visual, el orden del par es el inverso.

El método que se ha explicado es una heurística que se puede condensar en el código Pascal que aparece en el algoritmo 1.

### 10.5.5. Pruebas

Se ha realizado un programa informático para investigar diferentes aspectos de las RRTTDD y los GGRTD. El programa se ha implementado en Object Pascal usando el entorno de desarrollo Delphi XE7. El programa está diseñado para simular que un agente tiene dos acciones entre las que elegir para girar 1 grado hacia la izquierda o 1 grado hacia la derecha. El programa permite tres opciones en un espacio bidimensional que contiene un objeto y un agente. La forma y tamaño del objeto pueden ser modificadas. La primera opción, denominada en el menú *Manual Control*, permite al usuario controlar un agente usando un botón para girar a la izquierda y otro para girar a la derecha. El programa calcula la RTD entre el objeto y el agente en tiempo real. La segunda opción, denominada *Automatic Control*, permite al usuario determinar una RTD como objetivo, y el programa aplica la HDQS para establecer la RTD objetivo. La tercera opción, denominada *Generating DTRG*, genera el GRTD relativo al objeto y el agente mientras gira 360° al ir calculando cada RTR que se establece entre ambos. Cada opción crea dos pantallas: la vista aloctrica y la vista egocentrica. La vista aloctrica muestra las posiciones del objeto y del agente en tiempo real desde el pun-

**Algoritmo 1** Algoritmo de selección GRTD

---

**Entrada:** RTD :DirectionalTopologyR; Almacena la relación topológica direccional inicial  
id-GRTD:dtrgT;  
lado:integer; 1-Izquierda 2-Derecha

**Salida:** Identificador de grafo de razonamiento topológico direccional.

```

1: RTD :=calculandoRTD-Actual ;
   Calcula la RTD actual
2: IF ContrastarRTDClave(RTD)
   Examina si la RTD es una clave THEN
   BEGIN
     id-GRTD.caso := CalcularCaso(RTD);
     calcula el caso principal a partir de la RTD actual   ParClave:= EncontrarParClave;
     Gira para encontrar un par clave   id-GRTD.variante := CalcularVariacion(ParClave);
     Calcula la variación a partir del par clave   END
   ELSE
   BEGIN
     lado := SelectSpinning(DTR)
     Elige un lado hacia el que girar   WHILE NOT ContrastarRTDClave(DTR) DO
     Mientras no encuentre una clave girará.   BEGIN
       girarCuello(lado);
       RTD :=calculandoRTD-Actual ;
       END;
       id-GRTD.caso := CalcularCaso(RTD);
       calcula el caso principal a partir de la RTD encontrada.   WHILE anguloCuello < 0 DO
       Vuelve a la posición inicial   BEGIN
         girarCuello(Inverse(lado));
         ParClave:= CompruebaParClave;
         Gira para encontrar un par clave   END;
         id-GRTD.variante := CalcularVariacion(ParClave);
         Calcula la variación a partir del par clave   END
       result:= id-GRTD;

```

---

to de vista de un observador externo. La vista egocéntrica muestra las coordenadas con las que el agente percibe la localización del objeto. La vista egocéntrica también muestra la proyección del objeto en el eje horizontal mediante líneas rojas.

Durante las distintas pruebas se ha visto que hay ocasiones en las que el agente no detecta alguna RTR cuando está girando. Ese hecho se debe a que el agente gira una cantidad finita que se añade al ángulo de partida. Así, hay una secuencia de estados entre el estado de partida y el estado tras el giro. Si en alguno de los estados en esa secuencia hay una RTD diferente de la de partida a la que se tiene tras el giro, entonces esa RTD no es detectada. La RTD que solo es establecida con unas coordenadas específicas, por ejemplo, las RTD con su primer componente con *iguales*, *pegados-0* o *cubierto-0*, son susceptibles de no ser detectadas. Ese hecho se puede solucionar detectando cuándo el agente no establece una RTR por la falta de precisión. Esa detección se puede llevar a cabo detectando que no se ha llegado a la RTD objetivo, pero se ha alcanzado la RTR que sucede a la RTR objetivo en el grafo.

### 10.5.6. Sistema de coordenadas cualitativas esféricas

En el apartado 10.2.2 se vio que se podía hacer una generalización para aplicar la HSCT a espacios  $n$ -dimensionales; pero el espacio y los objetos que hay en el espacio deben de cumplir ciertas restricciones para que la heurística funcione. En la sección 10.4 se ha desarrollado una arquitectura capaz de navegar autónomamente en espacios desconocidos o dinámicos, pero el espacio tiene que tener las limitaciones que requiere la HSCT y además no se pueden tomar decisiones de giro. En esta sección se ha encontrado una heurística que permite tomar decisiones sobre giros. Las limitaciones sobre el espacio provienen de que la generalización para describir la relación espacial entre dos objetos se hace usando el producto cartesiano sobre el conjunto de las RRTTRR. Así, mientras que en un espacio de cualquier dimensión el conjunto de relaciones topológicas clásicas es siempre 8, para que la HSCT sea efectiva el número de relaciones se incrementa realizando el producto cartesiano por cada dimensión. Como se mencionó, las relaciones topológicas clásicas no permiten tomar decisiones de navegación, pero la generalización usando las RRTTRR necesita de muchas restricciones. Eso lleva a necesitar investigar nuevos métodos para representar las relaciones espaciales entre un objeto y un agente que necesiten menos restricciones que la generalización de las RRTTRR. En esa línea, mediante los resultados obtenidos en esta sección se propone el siguiente conjunto de relaciones para un espacio bidimensional:

$$D_{39} \times R_8$$

Así, la relación entre un objeto y el agente quedaría descrita por un par formado por una relación topológica direccional y una relación topológica clásica. Este par puede ser visto como las coordenadas cualitativas de un objeto en un sistema de referencia egocéntrico centrado en el agente. Por lo tanto, se denominará sistema de coordenadas cualitativas esféricas.

## 10.6. Generación de lenguaje natural

Una de los comportamientos más sobresaliente de los seres humanos es el de comunicarnos información por medio del lenguaje. Integrar robots, o agentes computacionales, en nuestra vida diaria no puede ocurrir sin el desarrollo de técnicas de lenguaje natural (LN). El LN es estudiado por diversas disciplinas, como la filosofía de lenguaje, la lingüística computacional, y la IA.

Los interlocutores de LN expresan principalmente relaciones cualitativas entre zonas espaciales u objetos para comunicar tareas o información. Por este motivo, el desarrollo de métodos que puedan generar y procesar información cualitativa es esencial para crear robots multifuncionales capaces de interacción continua en entornos sociales. La investigación presentada en este capítulo ha estado enfocada en permitir a los agentes

realizar diferentes tareas autónomas en un entorno a través de relaciones espaciales cualitativas. En esta sección, se presenta y discute una nueva etapa del programa RMMNT que tiene como objetivo generar y procesar LN concerniente a relaciones espaciales cualitativas. Esta sección presenta un programa para desarrollar esta nueva etapa y sus primeros resultados. La investigación realizada aquí está basada en la TCCV [371] que se presentó en el capítulo 9. La TCCV considera que las condiciones de verdad no son solo una herramienta para dar significado a expresiones lingüísticas, sino que juegan un papel fundamental en todas las capacidades cognitivas que los seres humanos poseen [371]. El resultado principal de esta sección es desarrollar un método para permitir a agentes autónomos generar expresiones simples de lenguaje natural sobre relaciones espaciales cualitativas.

Esta sección está estructurada de la siguiente manera: El apartado 10.6.1 revisa brevemente la investigación hasta la fecha en la generación y procesamiento de LN. El apartado 10.6.2 examina el problema del anclaje del símbolo y su importancia en la generación y procesamiento de lenguaje natural. El apartado 10.6.3 contiene el programa de la nueva etapa del RMMNT. El apartado 10.6.4 presenta un método nuevo para permitir a un agente autónomo generar expresiones de LN. Ese nuevo método implica la incorporación de nociones topológicas en la TCCV.

### 10.6.1. Planteamientos para el lenguaje natural

Dentro del LN se encuentran dos procesos que pueden ser estudiados: la generación y el procesamiento de lenguaje natural. Como se ha mencionado, hay diversas disciplinas que están interesadas en el fenómeno del LN. Así, existen diferentes planteamientos para describir y reproducir la generación y el procesamiento de LN, no solo por la existencia de diferentes puntos de vista; sino porque las metas no son las mismas. Respecto a el procesamiento de lenguaje natural, el objetivo de lingüistas, lógicos y filósofos son muy diferentes de los de los investigadores de IA y robótica. Mientras que los lingüistas, los lógicos y los filósofos intentan proporcionar teorías generales que determinen el significado de expresiones de lenguaje natural, investigadores de robótica tienden a enfocarse en desarrollar robots que producen un comportamiento adecuado a un conjunto de instrucciones en lenguaje natural.

Las teorías de la correspondencia, uno de los tipos de teorías generales del significado, explican las expresiones LN a través de métodos que conectan la expresión a elementos de una realidad, o modelo. Una directriz para construir esos métodos es el principio de composicionalidad según el cual *El sentido de una proposición es función del sentido de sus partes*. Uno de los tipos de modelos más potentes son los modelos de Kripke, que son conjuntos de mundos con una relación de accesibilidad entre los mundos. Cada mundo contiene una estructura que determina la granularidad que pueden tener las partes de la expresión denotando elementos del modelo. El uso de modelos semánticos de la lógica en la descripción de los significados de expresiones de lengua-

je naturales fueron desarrolladas inicialmente por Richard Montague, Maxwell John Cresswell, Renate Bartsch, y David Lewis, entre otros [171]. Sus trabajos mostraron que los métodos matemáticos pueden ser usados para explicar cómo los humanos manipulan el significado cuando ellos usan el lenguaje natural. Estos tipos de métodos han tenido como resultado la aparición del campo de la semántica formal[625, 217].

Las semánticas formales son usadas en IA cuando el modelo de expresiones de LN son datos organizados en una base de datos. La razón es porque las expresiones del LN pueden ser interpretadas como relaciones entre datos, y esto podría ser conocido si ellos son verdaderos o falsos inspeccionando la base de datos. Aunque las semánticas formales han sido muy exitosas en lingüística y programas de IA cuyo mundo es un entorno de software, no se pueden aplicar a la robótica debido al problema del anclaje del símbolo, que se analizará en el siguiente apartado. Para abordar esta problemática, los investigadores de IA han desarrollado técnicas alternativas para permitir que sus robots procesen LN. Una de las tareas más investigadas en robótica es hacer que un robot llegue a una localización espacial concreta siguiendo un conjunto de instrucciones en lenguaje natural dadas por un ser humano. Algunos métodos usados son *cláusulas de descripción espacial*[273], *campos semánticos*[146], y  $G^3$ [565]. Esos métodos no tratan de proporcionar una solución general para el procesamiento del lenguaje natural, sino que solo buscan manejar un número finito de instrucciones de LN del ámbito espacial. Estos métodos conectan la expresión con una secuencia de acciones que el robot ejecutará. Así, no necesitan cumplir el principio de composicionalidad porque no es necesario para los objetivos de los investigadores de robótica. Por ejemplo, el método de las cláusulas de descripciones espaciales no cumple el principio de composicionalidad, pero sin embargo es una técnica efectiva para las tareas que se han propuesto los investigadores.

### 10.6.2. El problema del anclaje del símbolo

En el capítulo 9, ya se mencionó que Tarski reintrodujo la noción de verdad en las matemáticas cuando dio una definición formal de verdad para los lenguajes formales, y que de esa noción formal de verdad surgió la rama de las matemáticas denominada teoría de modelos que estudia cómo los lenguajes formales describen clases de estructuras matemáticas. También se mencionó que el trabajo de Tarski fue la base sobre la que Davidson hizo su propuesta para explicar el significado del LN. La propuesta de Davidson ha sido desarrollada en las semánticas formales. Como se mencionó en el apartado anterior, si las expresiones de LN denotan relaciones en las bases de datos, o información almacenada en otras estructuras de datos, las semánticas formales pueden ser usadas para crear un programa que procese expresiones de LN y actúe de acuerdo en un entorno software. Sin embargo, estas técnicas no pueden ser usadas si el modelo no es dado en una forma accesible. La problemática puede ilustrarse usando el verbo “mover”. Las semánticas formales dicen que el verbo “mover” se puede traducir como

un predicado  $M(x)$  o como la función  $\lambda x.M(x)$ . Siguiendo la propuesta de Davidson, un esquema T determinará si una expresión es verdad o falsa verificando si lo que la expresión denota existe o no en el modelo. Por lo tanto, el significado de  $M(x)$  es el conjunto  $\mathbf{M}$  en el modelo. Se usa *itálica* para denotar elementos del lenguaje y **negrita** para denotar elementos del modelo. Así, la expresión “Juan se mueve” se traduce por  $M(\textit{Juan})$ , la cual, es verdad si y solo si **Juan**  $\in$   $\mathbf{M}$ . Un lingüista que usa semánticas formales siempre puede acceder al modelo para verificar porque él ha definido el modelo y la interpretación, y, por lo tanto, conoce la interpretación. El problema surge cuando no se conoce el modelo ni la interpretación porque el acceso al modelo no es sencillo. Esto ocurre cuando el modelo es el mundo, lo que ocurre en el caso de la robótica. Ni Davidson, ni las semánticas formales, han explicado cómo se puede decidir qué objeto, o propiedad de un objeto del mundo, encaja con un símbolo. Si se tiene un método que permite determinar si un objeto es denotado o no por un símbolo, entonces ese símbolo está anclado. La cuestión de cómo lograr métodos para establecer que los símbolos de un lenguaje estén anclados es denominado, el problema del anclaje del símbolo [209, 543]. Las semánticas formales son una herramienta potente para manejar LN pero sin un método para permitir que el agente robótico pueda verificar el mundo autónomamente, no se pueden usar en robótica. Ante esta situación, ha existido un esfuerzo en investigar el problema del anclaje del símbolo en robótica. La investigación ha mostrado que es posible para un robot crear símbolos anclados [543], pero esos resultados han mostrado solo la existencia de métodos para expresiones específicas como los colores. Así, se está lejos de tener un método que permita el uso de semánticas formales en robótica. Por lo tanto, abordar métodos de anclaje de símbolos para la generación y procesamiento de lenguaje natural es una cuestión crítica para el avance de la robótica, así como para comprender cómo los exocomportamientos de los seres humanos relacionados con el LN emergen.

### 10.6.3. Objetivos y planificación de la nueva etapa

La nueva etapa del programa RMMNT investiga cómo las nociones topológicas pueden ser usadas en la generación y procesamiento de LN. El principal objetivo de la nueva etapa es desarrollar métodos para expresar y reconocer expresiones de LN sobre relaciones espaciales usando nociones topológicas, aunque este no es el único objetivo. Esta nueva etapa también persigue métodos para implementar una teoría general de LN porque si se enfocara en solo una tarea concreta, no se pueden desarrollar robots multifuncionales. Se considerará que un método implementa una teoría general cuando cumpla el principio de composicionalidad. Junto con esos objetivos, será necesario abordar el problema del anclaje del símbolo, y específicamente, la creación de métodos para el anclaje de verbos. Para hacer esto, la investigación partirá de la TCCV que ha sido presentada en el capítulo 9 porque es una propuesta que está enfocada específicamente en el anclaje de verbos.

Durante esta nueva etapa, el programa RMMNT asume que “la señal de voz” ha sido preprocesada como una cadena de palabras o que puede ser postprocesada de una cadena de palabras a señales de voz. Por ello, los procesos de preprocesamiento y postprocesamiento quedan fuera de la investigación que se ha planificado. La planificación de la investigación que se ha propuesto para la segunda fase del programa RMMNT es la siguiente:

1. Generar expresiones simples de lenguaje natural.
  - a) Producir frases verbales relacionadas con información espacial del pasado y el presente.
  - b) Producir expresiones con cuantificadores (ej. un, dos, varios,...).
  - c) Producir frases verbales relacionadas con información espacial del futuro.
2. Procesamiento del lenguaje natural.
  - a) Procesar expresiones declarativas simples relacionadas con información espacial. La meta es que el agente determine si la afirmación simple con información espacial cualitativa es verdadera o falsa.
  - b) Procesar los actos de habla: solicitar (metas directas espaciales) y etiquetar objetos. Este punto de la planificación pretende abordar el comienzo de la teoría de actos de habla [127].
3. Conversación.
  - a) Añadir al procesamiento nuevos actos de habla: llamar, saludar, y protestar.
  - b) Completar la generación de los restantes actos de habla.
4. Abordar expresiones complejas de lenguaje natural ( ej. subordinadas, coordinadas,...) .

Un lector podría preguntarse por qué, si la mayoría de las investigaciones de robótica comienzan abordando el procesamiento de expresiones de lenguaje natural, está planificación comienza con la generación de expresiones de LN. Obviamente, un robot solo puede colaborar con humanos no expertos cuando puede procesar LN, y un robot que solo puede generar expresiones de LN no es útil porque no puede recibir instrucciones. Sin embargo, hay una razón importante para escoger este camino aparentemente poco provechoso para investigar LN. Esta planificación permite el desarrollo de métodos que cumplan con el principio de composicionalidad porque se evita el tener que ser efectivo en una tarea concreta, ya que en la generación de LN no se tiene ninguna meta para hacer que el agente robótico responda realizando algún comportamiento concreto. Por lo tanto, una vez se desarrollen métodos para generar LN que cumplen el principio de composicionalidad se puede abordar, con el conocimiento adquirido, el desarrollo de métodos para el procesamiento de LN que cumplan el LN.

#### 10.6.4. Generando lenguaje natural mediante condiciones de verdad

La investigación que se presenta en esta sección ha estado enfocada en crear un método para generar sentencias simples de LN. El núcleo de una sentencia simple es un verbo, que estructura la sentencia. Durante los primeros pasos de la investigación, el verbo *estar* ha sido elegido porque permite describir relaciones espaciales, ej. Juan está en el Empire State Building. La meta específica que se ha abordado ha sido anclar el verbo *estar*. A continuación, se hace una descripción general de cómo funciona el método que se ha hallado en la investigación.

El primer requisito para generar una expresión en LN es que el agente robótico debe tener una fórmula para describir lo que ha ocurrido en un intervalo de tiempo. Sin embargo, esto no puede suceder directamente porque un agente robótico solo puede tener acceso al presente. Para afrontar esa problemática, la investigación se ha basado en la TCCV porque su jerarquía de lenguajes replica el proceso de generación de información cualitativa en el nivel simbólico cómo ocurre en el sistema nervioso. La TCCV resuelve la cuestión de que el agente solo tiene acceso a la información del presente en base al hecho físico de que el tiempo fluye constantemente y el agente está constantemente cambiando a una nueva perspectiva temporal. Así, la TCCV propone que el agente cree la denominada *formula historial*. La creación de la formula historial comienza transformando la descripción del presente,  $\phi$ , que ha sido registrada por los sensores en una descripción situada en el pasado  $\phi'$ , y con la siguiente lectura de sus sensores que haga el agente robótico sustituirá la fórmula del presente, y almacena la versión situada en el pasado. Este proceso se repetiría continuamente. Así, cuando  $\phi'$  no es la primera en almacenarse, entonces se usa el conector  $\blacktriangleright$  para concatenar  $\phi'$  con la fórmula almacenada  $\varphi$ , produciendo la fórmula  $\varphi \blacktriangleright \phi'$ . Mediante este proceso el agente construye una formula  $L_{PPFMM}$  que describe lo que ha ocurrido del pasado al presente.

La TCCV tiene como base de la jerarquía de lenguajes a  $L_{PPFMM}$ , que está construida con la información registrada directamente por sensores. Así, usándose la fórmula historial de  $L_{MMPPF}$ , puede generar una fórmula historial de  $L_{MMPPF}^*$ , y usar esta última para generar una fórmula historial de  $LC_{PPFMM}$ . Sin embargo, hay una cuestión que debe ser abordada sobre los lenguajes que propone la TCCV para cumplir con los objetivos del programa RMMNT. Esa cuestión es que los lenguajes de la TCCV no están definidos para expresar relaciones espaciales. Por lo tanto, los alfabetos de los lenguajes de la TCCV deben ser expandidos.

Una vez el agente tiene una fórmula historial  $LC_{PPFMM}$ , la cuestión es cómo generar LN a partir de ella. Los lenguajes de la jerarquía son solo descripciones privada que posee un agente, y LN es una descripción que es liberada al entorno del agente. Por lo tanto, el principal problema a abordar es como transformar la descripción privada en la descripción pública. Dado que una de las metas es cumplir con el principio de

composicionalidad, la idea general bajo el método que se ha desarrollado es que la fórmula historial  $LC_{PPFMM}$  debe ser rota en partes, de manera que lo que denote la fórmula sea función de lo que denotan la partes de la fórmula, y que esas partes puedan ser traducidas en palabras.

La idea general que se ha expuesto puede ser llevada a cabo debido al hecho de que las fórmulas de  $LC_{PPFMM}$  pueden ser construidas mediante  $\beta$ -reducción a secuencias de átomos de las categorías de  $LC_{PPFMM}$ . Así, si se recuperan los átomos que componen una fórmula de  $LC_{PPFMM}$ , esos átomos son las partes que se pueden transformar directamente en palabras con las que componer la expresión LN que el agente debe generar. Los átomos pueden ser recuperados aplicando la operación inversa a la  $\beta$ -reducción a una fórmula  $LC_{PPFMM}$ . Esa operación está definida en el  $\lambda$ -calculus, la  $\beta$ -abstracción. Así, la relación inversa entre  $\beta$ -reducción y  $\beta$ -abstracción nos indica que la operación de  $\beta$ -reducción va a estar relacionada con el procesamiento de LN y la de  $\beta$ -abstracción con la generación de LN.

Obtener los átomos no es el final del proceso. Hay dos procesos que deben ser completados para producir expresiones de LN: Traducir átomos en palabras y ordenar esas palabras. La traducción de átomos en palabras es un proceso importante para el que el agente tiene que tener estructuras de datos, a las que se denominarán: estructuras de producción de lenguaje público (EEPLP). Una EPLP almacena átomos de una categoría, o una fórmula cuasiatómica generada por la reducción de átomos, y estos están asociados con cadenas de caracteres. La ordenación de las palabras sería llevada a cabo mediante gramáticas generativas. A continuación, se describirán en detalle los pasos de generación de expresiones de LN con el verbo *estar*.

#### 10.6.4.1. Expandiendo $L_{PPFMM}^*$ y $CL_{PPFMM}$

Como se ha mencionado, los lenguajes definidos inicialmente en la TCCV deben ser expandidos para permitir que describan relaciones espaciales, ya que las definiciones de los alfabetos de metainformación de  $L_{PPFMM}^*$  solo cualifican el cambio de valores con dominios que tienen relaciones totales de orden [371]. Así, si la posición de los objetos en el modelo no es descrito con un solo punto del espacio, hay relaciones espaciales que no pueden ser descritas por  $L_{PPFMM}^*$ . Dado que la TCCV tiene una jerarquía de lenguajes, si se modifica uno de ellos, todos los lenguajes sobre ese tendrán que ser modificados también. En este caso el lenguaje  $L_{PPFMM}$  no necesita ninguna modificación porque los sensores de un agente robótico no registran relaciones cualitativa. Puesto que  $L_{PPFMM}^*$  construye información cualitativa de  $L_{PPFMM}$ , el lenguaje  $L_{PPFMM}^*$  debe ser expandido<sup>5</sup> para describir relaciones espaciales cualitativas. Debe tenerse en cuenta que la extensión de un lenguaje de la jerarquía debe de realizarse usando el lenguaje

<sup>5</sup>Se podría crear una jerarquía dentro del propio  $L_{PPFMM}^*$ , ya que se puede describir si una relación topológica está cambiando o no; pero no abordaremos este tema aquí.

que existe en el nivel anterior para cumplir con el principio de la jerarquía de lenguajes de la TCCV.

El primer paso es añadir un nuevo alfabeto de metainformación a  $L_{PPFMM}^*$ ,  $\Sigma_T$ , cuyos elementos describen relaciones topológicas entre dos objetos y un nuevo tipo de fórmula atómica que usa este nuevo alfabeto de metainformación. En esta investigación se ha considerado que  $\Sigma_T$  tiene las relaciones relativas topológicas. Las relaciones espaciales relativas son relaciones en un espacio unidimensional, pero pueden ser usadas para espacios de mayores dimensiones como ya se ha mencionado.

El nuevo tipo de fórmula atómica es:

$$[\boxtimes|tp|o_x|o_y|0|\sigma_u] \quad \boxtimes \in \{h, \varepsilon\} \quad tp \in \{E_{\downarrow}, @_{\parallel}, E_{\uparrow}\} \quad o_x, o_y \in \mathcal{V} \quad \sigma_u \in \Sigma_T$$

El valor 0 en el quinto campo del nuevo tipo de fórmula determina que la relación que la fórmula está describiendo una relación espacial ya que el 0 identifica la propiedad del espacio.

Este nuevo tipo de fórmula atómica de  $L_{PPFMM}^*$  es construida a partir de  $L_{PPFMM}$  definiendo una función, *ett*, denominada *estado topológico temporal* con la siguiente definición:

$$\begin{aligned} tts : \mathcal{P}(V_0) \times \mathcal{P}(V_0) &\longrightarrow \Sigma_T \\ (s_x, s_y) &\longmapsto \sigma_v \end{aligned}$$

donde  $V_0$  es el dominio cuyos elementos son puntos del espacio y la definición de *tts* es hecha de acuerdo a las definiciones dadas de las relaciones relativas espaciales. Por lo tanto, una fórmula topológica atómica de  $L_{PPFMM}^*$  es construida de la siguiente manera:

$$[\boxtimes|tp|0|o_x|o_y|0|tts(g^{*0}(o_x), g^{*0}(o_y))] \quad \boxtimes \in \{h, \varepsilon\} \quad tp \in \mathcal{C}_{PT} \quad o_x, o_y \in \mathcal{C}_O \quad \sigma_u \in \Sigma_T$$

donde la función  $g^{*0}$  determina el espacio ocupado por un objeto en  $L_{PPFMM}$ . Estas nuevas fórmulas se combinan con las fórmulas ya definidas con los conectores que contiene  $L_{PPFMM}$ .

La extensión de  $L_{MMPPF}^*$  permite extender también  $LC_{PPFMM}$ . Así, se añade  $\Sigma_T$  a los alfabetos de metainformación de  $LC_{PPFMM}$ , y la siguiente fórmula atómica:

$$[\boxtimes|tp|o_x|o_y|0|\sigma_u]$$

La generación de este nuevo tipo de fórmula de  $LC_{PPFMM}$  se puede generar a partir de una fórmula historial de  $L_{PPFMM}^*$  mediante una gramática reconocedora de condiciones de verdad como se ha explicado en el capítulo 9.

### 10.6.4.2. La extracción de átomos de fórmulas de $CL_{MMPPF}$

Dado que el nuevo tipo de fórmula atómica mostrada en el apartado anterior es añadida a  $LC_{PPFMM}$ , también el siguiente tipo de átomo debe ser añadido a la categoría de registros de estados.

$$\lambda \boxtimes \lambda t p \lambda o_x \lambda o_y \lambda \sigma_u [\boxtimes | t p | o_x | o_y | 0 | \sigma_u ]$$

Este átomo es equivalente al significado del verbo *estar* cuando describe relaciones cualitativas espaciales entre dos objetos. Nótese el hecho de que el átomo es básicamente una estructura sintáctica que tiene la semántica de que hay una relación espacial entre dos objetos.

Como se ha mencionado en la descripción general del método, a partir de una fórmula de  $CL_{MMPPF}$  los átomos deben ser extraídos a través de  $\beta$ -abstracción. Se aplica  $\beta$ -abstracción hasta que todos los átomos hayan sido extraídos del registro de estado. Por ejemplo, si se tiene la fórmula  $[\varepsilon | | | o_5 | o_7 | 0 | \sigma_u ]$ , que describe que el objeto  $o_5$  está dentro del objeto  $o_7$ , y se aplica  $\beta$ -abstracción repetidamente se obtienen seis átomos.

$$[\varepsilon | | | 0 | o_5 | o_7 | 0 | \sigma_u ] \leftarrow_{\beta} \lambda \sigma_u [\varepsilon | | | o_5 | o_7 | 0 | \sigma_u ] s_7$$

$$\lambda \sigma_u [\varepsilon | | | o_5 | o_7 | 0 | \sigma_u ] s_7 \leftarrow_{\beta} \lambda o_y \lambda \sigma_u [\varepsilon | | | o_5 | o_y | 0 | \sigma_u ] (o_7) (s_7)$$

$$\lambda o_y \lambda \sigma_u [\varepsilon | | | o_5 | o_y | 0 | \sigma_u ] (o_7) (s_7) \leftarrow_{\beta} \lambda o_x \lambda o_y \lambda \sigma_u [\varepsilon | | | o_x | o_y | 0 | \sigma_u ] (o_5) (o_7) (s_7)$$

$$\lambda o_x \lambda o_y \lambda \sigma_u [\varepsilon | | | o_x | o_y | 0 | \sigma_u ] (o_5) (o_7) (s_7) \leftarrow_{\beta} \lambda t p \lambda o_x \lambda o_y \lambda \sigma_u [\varepsilon | t p | o_x | o_y | 0 | \sigma_u ] (| | |) (o_5) (o_7) (s_7)$$

$$\begin{aligned} \lambda t p \lambda o_x \lambda o_y \lambda \sigma_u [\varepsilon | t p | o_x | o_y | 0 | \sigma_u ] (| | |) (o_5) (o_7) (s_7) & \leftarrow_{\beta} \\ \leftarrow_{\beta} \lambda \boxtimes \lambda t p \lambda o_x \lambda o_y \lambda \sigma_u [\boxtimes | t p | o_x | o_y | 0 | \sigma_u ] (\varepsilon) (| | |) (o_5) (o_7) (s_7) \end{aligned}$$

### 10.6.4.3. El vector de roles-átomos

Una vez los átomos son obtenidos, es necesario almacenarlos en una estructura de datos denominada vector de roles-átomos. La razón de ese nombre quedará clara a continuación. El vector contiene un primer elemento que es siempre un átomo de la clase  $\mathcal{C}_{RE}$ , al que se llamará núcleo. Los elementos siguientes al núcleo del vector son pares que tienen como primer elemento del par, la variable generada en la  $\beta$ -abstracción, y el átomo que se extrajo como segundo elemento. Así, continuando con el ejemplo anterior, su vector de roles-átomos es el siguiente:

$$\langle \lambda \mathfrak{X} \lambda t p \lambda o_x \lambda o_y \lambda \sigma_u [\mathfrak{X} | t p | o_x | o_y | 0 | \sigma_u], (\mathfrak{X}, \varepsilon), (t p, ||), (o_x, o_5), (o_y, o_7), (\sigma_u, s_7) \rangle$$

#### 10.6.4.4. Estructuras de producción de lenguaje público

Previamente, se mencionó la necesidad de una EPLP en el proceso de generación de expresiones de LN. Una opción fácil de crear una EPLP es una tabla, pero perfectamente se podría usar otra estructura de datos que permitiera una recuperación más rápida de información. La tabla asociada con la categoría registros de estado tiene dos campos en cada registro. Un campo contiene el átomo y el otro una cadena de caracteres, denominada v-cadena, que corresponde con una palabra de LN. Las tablas 10.11 y 10.12 muestran cómo los núcleos son traducidos a LN. Las tablas de los núcleos tienen dos niveles. El primer nivel, que muestra la tabla 10.11, contiene los núcleos y sus correspondientes verbos. El segundo nivel, que muestra la tabla 10.12 contiene fórmulas cuasiatómicas y las formas del modo de cada verbo. Las fórmulas cuasiatómicas del segundo nivel son generadas mediante  $\beta$ -reducción, pero la  $\beta$ -reducción solo se realiza con posicionadores temporales y con las condiciones de realidad. Más adelante se mostrará un ejemplo.

Átomo	v-cadena	Id. de la tabla de segundo nivel
$\lambda \mathfrak{X} \lambda t p \lambda o_x \lambda o_y \lambda \sigma_u [\mathfrak{X}   t p   o_x   o_y   0   \sigma_u]$	estar	5
$\lambda \mathfrak{X} \lambda t p \lambda o_x [\mathfrak{X}   t p   o_x   0   \beta_2]$	mover	10

Tabla 10.11: EPLP de núcleos

Identificador de la tabla: 5	
Átomo	v-cadena
$\lambda o_x \lambda o_y \lambda \sigma_u [\varepsilon        o_x   o_y   0   \sigma_u]$	está
$\lambda o_x \lambda o_y \lambda \sigma_u [\varepsilon   \downarrow     o_x   o_y   0   \sigma_u]$	estaba

Tabla 10.12: Tabla del segundo nivel de la EPLP de los núcleos asociada con el verbo *estar*

Se crean también tablas para las relaciones espaciales cualitativas y el resto de tipos de átomos. Debe notarse que los identificadores de objetos son nombres propios.

### 10.6.4.5. La gramática generativa

El orden de la  $\beta$ -reducción es importante porque determina en qué variables se hace la sustitución, lo que implica que el orden de emisión de palabras es importante de cara a la creación de métodos para el procesamiento de LN. Además, experimentalmente se sabe que cambiar el orden de las palabras implica cambios en el significado. Por lo tanto, el orden en el que las palabras son emitidas es importante. El proceso usa una gramática generativa para determinar el orden en el cual las palabras son emitidas. Los símbolos no terminales de la gramática son variables generadas para la  $\beta$ -abstracción.

$$\begin{array}{ll}
S & \longrightarrow N \\
N & \longrightarrow MVRO \\
V & \longrightarrow \lambda \mathfrak{X} \lambda tp \lambda o_x \lambda o_y \lambda \sigma_u [\mathfrak{X} | tp | o_x | o_y | 0 | \sigma_u ] (\mathfrak{X})(tp) \\
M & \longrightarrow o_x \\
O & \longrightarrow o_y \\
R & \longrightarrow \sigma_u
\end{array}$$

Los elementos terminales producidos por la gramática son emparejados dentro de las EEPLP para extraer las v-cadenas. El proceso de producción tiene un caso especial al aplicar las reglas. Ese caso es cuando la regla produce un  $\lambda$ -término. En este caso hay que obtener términos del vector de roles-átomos y aplicar  $\beta$ -reducción. Por ejemplo, tras aplicar la regla para  $V$  la siguiente expresión es obtenida:

$$\lambda \mathfrak{X} \lambda tp \lambda o_x \lambda o_y \lambda \sigma_u [\mathfrak{X} | tp | o_x | o_y | 0 | \sigma_u ] (\mathfrak{X})(tp)$$

junto con el  $\lambda$ -término se ha generado dos variables de términos. Esas variables tienen que ser substituidas por términos. Así, usando el vector de roles-átomos obtenido previamente,

$$\langle \lambda \mathfrak{X} \lambda tp \lambda o_x \lambda o_y \lambda \sigma_u [\mathfrak{X} | tp | o_x | o_y | 0 | \sigma_u ], (\mathfrak{X}, \varepsilon), (tp, ||), (o_x, o_5), (o_y, o_7), (\sigma_u, s_7) \rangle$$

se substituyen las variables de términos, generando la siguiente  $\lambda$ -expresión:

$$\lambda \mathfrak{X} \lambda tp \lambda o_x \lambda o_y \lambda \sigma_u [\mathfrak{X} | tp | o_x | o_y | 0 | \sigma_u ] (\varepsilon)(||)$$

Una vez se tiene la anterior fórmula, se aplica  $\beta$ -reducción, lo que genera la siguiente fórmula:

$$\lambda o_x \lambda o_y \lambda \sigma_u [\varepsilon | || | o_x | o_y | 0 | \sigma_u ]$$

Esta fórmula final es la que se empareja en la tabla 10.12 para obtener la v-cadena, que en este caso es *está*.

De cara al procesamiento de LN, hay que señalar que la flexibilidad en el orden de las palabras depende del uso de preposiciones que determinen unívocamente que variable será substituida en el proceso de  $\beta$ -reducción.

#### 10.6.4.6. Generación de lenguaje natural basado en condiciones de verdad

El algoritmo para generar una expresión de LN sobre dos objetos específicos y la relación espacial entre ellos que se propone se puede resumir en los siguientes pasos:

1. El agente registra información usando sus sensores para generar una fórmula bien formada del  $L_{PPFMM}$  y construye una fórmula historial.
2. El agente transforma la fórmula historial  $L_{PPFMM}$  en una fórmula  $L^*_{PPFMM}$
3. El agente usa una gramática reconocedora de condiciones de verdad para generar una fórmula bien formada del  $LC_{PPFMM}$  a partir de la fórmula  $L^*_{PPFMM}$  previa.
4. El agente usa abstracción  $\beta$  para extraer los átomos de la fórmula  $CL_{PPFMM}$  generada en el paso anterior.
5. El agente genera el vector de roles-átomos.
6. El agente usa una gramática generativa para crear una estructura sintáctica de la sentencia de LN.
7. El agente aplica reducción  $\beta$  si fuera necesario.
8. El agente recupera las v-cadena de las estructuras de producción de lenguaje público y sustituye los átomos por las v-cadena en la estructura sintáctica.

### 10.7. Análisis de los resultados

Este capítulo ha presentado los resultados obtenidos en el programa RMMNT. Este programa se ha propuesto para abordar el objetivo de contribuir al desarrollo de métodos que permitan la creación robots multifuncionales. La primera etapa del programa RMMNT tiene como objetivo estudiar cómo las nociones topológicas pueden ser usadas para resolver el problema de navegación basado en relaciones espaciales.

Los primeros resultados fueron un estudio matemático profundo de la HSCT y su generalización para aplicarla en un espacio n-dimensional. Tras esos resultados se comenzó el estudio de extensiones de la HSCT, cuyos resultados se han presentado en la sección 10.3. Concretamente se investigaron dos posibles líneas de extensión.

La primera línea investigada para extender la HSCT ha sido añadir más relaciones al conjunto de las RRTTRR, ya que hay alguna situaciones espaciales que no pueden ser descritas por las relaciones de  $S_{13}$  si se permite que los objetos tengan agujeros. Concretamente se ha investigado la posibilidad de añadir relaciones de cercado. Las

relaciones que se definan no pueden provocar que el grafo que se cree para el razonamiento no sea lineal ya que eso impediría que se puedan tomar decisiones. La extensión propuesta permite describir las relaciones espaciales entre dos objetos en los que uno o los dos tienen al menos un agujero. Los resultados permiten concluir que el número de casos a contemplar para asignar un GLRT crece exponencialmente, lo llevaría un importante esfuerzo de programación para implementar métodos que usen las relaciones de cercamiento. Sin embargo, la investigación que se ha hecho en esta línea es solo un primer paso para entender como afrontar la toma de decisiones con objetos o regiones que tienen agujeros.

La segunda línea de investigación para extender la HSCT ha sido usar relaciones topológicas borrosas. El uso de reacciones topológicas borrosas es un tema que es investigado en el campo de los sistemas de información geográfica pero antes de esta investigación no se había considerado su uso en el campo de la robótica. Una razón es que los métodos de navegación basados en nociones topológicas apenas se están comenzando a investigar, por lo tanto, el uso de métodos topológicos borrosos es un campo inexplorado. La definición de relaciones topológicas borrosas propuesta en la investigación difiere del tipo de definiciones hechas en otras investigaciones [214, 560, 308, 58] porque en las investigaciones anteriores se parte de un espacio topológico borroso. Sin embargo, en la investigación que se ha llevado a cabo, las relaciones topológicas borrosas son definidas en un espacio topológico clásico. Una cuestión interesante de la investigación es que en el método desarrollado para la toma de decisiones combina el uso de relaciones clásicas y borrosas.

La HSCT es un método que solo permite la toma de decisiones para un objetivo concreto, mientras que la creación de robots multifuncionales implica dotar al robot de cierta autonomía que le permita autogenerarse metas. Además, un robot necesitaría ser capaz de navegar en espacios dinámicos o desconocidos para poder colaborar con los seres humanos en su vida cotidiana. Buscando abordar esos objetivos, en la investigación se ha diseñado la ANCT 1.0 que se ha presentado en la sección 10.4. La ANCT 1.0 no lleva a cabo planificación sino que reacciona al entorno. Sin embargo, aunque es una arquitectura reactiva difiere de algoritmos como BUG en que reacciona a relaciones espaciales y no a valores concretos detectados por los sensores. Esta característica permite dotar a la ANCT 1.0 con un alto grado de abstracción frente a los algoritmos reactivos que usan valores detectados por los sensores de distancia o contacto como disparadores de sus reacciones. Debido a este resultado, ANCT 1.0 es denominada una arquitectura de reactiva de segundo orden.

Otra cuestión importante de la ANCT 1.0 es que para un observador externo la arquitectura desarrolla tanto comportamientos reactivos como proactivos. Esto es logrado porque la arquitectura genera dos tipos de objetos en su análisis del espacio: objetos reales y objetos virtuales. Los objetos reales existen en el entorno y son detectados por los sensores. Estos objetos provocan el comportamiento reactivo que ve un observador externo. Los objetos virtuales no existen en el entorno, pero se generan

también a partir de la información que recogen los sensores. Estos objetos provocan el comportamiento proactivo que percibe el observador. El comportamiento de un agente que implementa la ANCT 1.0 está basado en el concepto de tarea topológica. Las tareas topológicas permiten a un agente tener suficiente autonomía para autogenerarse sus propias metas.

La ANCT 1.0 tiene la limitación de que la navegación debe ser en espacios abiertos que no requieran giros. Así, se ha investigado la toma de decisiones sobre giros en un espacio bidimensional. En esa investigación se ha desarrollado la HSCD, una heurística que permite tomar decisiones para establecer relaciones direccionales entre un objeto y un agente. Para desarrollar la HSCD, se ha creado un método para codificar información direccional mediante relaciones espaciales. El método de codificación tiene dos pasos. Primero, se asigna una región al objeto y al agente en un espacio unidimensional. Tras eso, se calcula la RTR de las dos regiones en el espacio unidimensional. El segundo paso es calcular desde que región del espacio se hace la proyección al espacio unidimensional para asignar una variante posicional que evitará la pérdida de información por la proyección del espacio bidimensional al espacio unidimensional. La propuesta desarrollada, propone el uso de 3 variantes posicionales. La combinación de las RRTTRR con las variantes posicionales crea las RRTTDD. Existen 39 RRTTDD, y se debe notar que esas relaciones son únicamente información direccional, aunque esté codificada usando RRTTRR. Además de las RRTTDD para la representación de estados, la HSCD hace uso de los GGRTD, que son grafos cíclicos que organizan una representación de estados de un agente y un objeto mediante RRTTDD. La HSCD requiere que el agente y el objeto no colisionen mientras el agente gire y el objeto no puede cercar al agente, ya que en esa situación no hay información direccional al estar el objeto en todas las direcciones respecto al agente.

La HSCD es un resultado importante para abordar el uso de la ANCT 1.0 en espacios cerrados, y el uso de robots comerciales, ya que muchos robots comerciales entre su repertorio de acciones básicas están los giros y no los desplazamientos laterales. La investigación sobre giros puede continuar abordando la predicción de colisiones. Además, también se puede explorar el uso de un número mayor de variantes posicionales para determinar la procedencia de las proyecciones, y estudiar conjuntos diferentes de RRTTDD formados a partir de conjuntos con diferente número de variantes posicionales.

Una cuestión importante que debe notarse de la investigación es que las acciones de giro no son las únicas que modifican la RTD entre un agente y un objeto. Las acciones de avanzar, retroceder, desplazarse a la izquierda y derecha también afectan a la RTD. Además, hay que tener en cuenta que en los GRTD no están todas las RRTTDD de  $D_{39}$ , y que los que no se usan podrían ser usados para tomar decisiones sobre desplazarse a la izquierda derecha, avanzar y retroceder.

En base a los resultados obtenidos se ha propuesto combinar las RRTTDD con las relaciones topológicas clásicas para definir unas coordenadas cualitativas que eviten

las limitaciones que tienen las RRTTRR en espacios de más de una dimensión. A las anteriores coordenadas se las ha denominado coordenadas cualitativas esféricas. Estas coordenadas abren una línea de investigación para combinar información direccional y topológica cualitativa para la toma de decisiones.

En base a todo lo anterior, se puede concluir que los resultados obtenidos hasta el momento en la primera etapa de la investigación del RMMNT son positivos y existen numerosas cuestiones para continuar investigando.

La última sección del capítulo ha abordado el comienzo de una nueva etapa del programa RMMNT, enfocada en producir y procesar expresiones de LN referentes a relaciones espaciales cualitativas. El comienzo de esta nueva etapa no significa que la etapa previa está terminada. El análisis ha mencionado múltiples cuestiones para continuar investigando en la primera etapa del RMMNT. Sin embargo, el conjunto de resultados obtenidos en la primera etapa son un conjunto de resultados lo suficientemente amplio para abordar la generación y procesamientos de LN para un robot multifuncional. Así, se ha presentado una planificación para la nueva etapa y sus primeros resultados. La segunda etapa no busca lograr métodos para que los humanos puedan dar instrucciones a un agente robótico a corto plazo, sino hacer una investigación profunda sobre LN. Así, los primeros pasos presentados aquí han sido enfocados en diseñar un método para resolver el problema del anclaje del símbolo mientras se cumple el principio de composicionalidad. El método que se ha presentado está basado en la TCCV, aunque la necesidad de tratar con relaciones espaciales cualitativas ha llevado a la expansión de los lenguajes de la TCCV. El método descrito ha sido aplicado a la generación de sentencias simples sobre relaciones espaciales entre dos objetos con el verbo *ser* como núcleo de sentencia. El método presentado aquí se ubica el punto 1.a de la planificación de la fase segunda. El siguiente paso es implementar este método en los agentes del micromundo en los que se implementó la ANQT 1.0 y anclar más verbos.

El método que se ha presentado está basado en la TCCV, pero la investigación realizada también ha contribuido a desarrollar la TCCV. Se debe prestar atención a cómo en la investigación surge una sinergia entre el programa RMMNT y la TCCV. La TCCV es muy útil en la nueva etapa del programa RMMNT pero además la nueva etapa ayudará a desarrollar la TCCV. Además, la investigación que ha comenzado en esta nueva etapa puede ayudar a desarrollar una teoría de la que se carece para explicar la interrelación entre las relaciones espaciales categóricas lingüísticamente y no lingüísticamente codificadas [263].



# Capítulo 11

## Conclusiones

En esta tesis doctoral se han presentado las investigaciones que se han llevado a cabo para formular y desarrollar una teoría de unificación de los comportamientos de sistemas biológicos y no biológicos autónomos. Al comienzo de la investigación se ha llevado a cabo una revisión general de las propuestas que se han realizado para explicar los comportamientos, la cual ha sido presentada en el capítulo 2. De dicha revisión se ha concluido que las nociones de inteligencia y cognición no pueden permitir la unificación de los comportamientos debido a incompatibilidades de esos conceptos con comportamientos descubiertos en la naturaleza. Además, ambos conceptos se encuentran en un estado de ambigüedad desde hace un largo periodo de tiempo, lo que hace razonable investigar nociones alternativas para explicar los comportamientos. A la anterior revisión hay que añadir otra que se ha realizado otra sobre los efectos de los comportamientos en los procesos evolutivos en el capítulo 3. Esta segunda revisión ha mostrado que los comportamientos no son fenómenos aislados; sino que son a la vez una de las causas de la dirección que toma la evolución. Por lo tanto, se ha concluido que una teoría de unificación de los comportamientos debe abordar cómo el comportamiento determina la dirección de la evolución. Así, dado que las teorías propuestas para explicar los comportamientos solo se han ocupado del propio comportamiento y la necesidad de investigar nociones alternativas a la de inteligencia, la investigación se ha centrado en la teoría general del exocomportamiento (TGE) como teoría para lograr unificar los comportamientos. La TGE es una propuesta muy reciente por lo que primeramente se ha llevado a cabo un trabajo para obtener una formulación más precisa de la que se tenía. La nueva formulación se ha presentado en el capítulo 4 y permite diferenciar fenómenos que las formulaciones previas no lo hacían. La TGE define un conjunto de fenómenos físicos, denominado exocomportamientos. El conjunto de los exocomportamientos contiene los comportamientos de sistemas biológicos y no biológicos autónomos, pero además incluye los fenómenos de desarrollo estructurales de los sistemas biológicos. La TGE propone que la causa de los exocomportamientos es una propiedad denominada fasa. La formulación de la fasa cumple con la hipótesis

del fisicalismo de los comportamientos expuesta en el capítulo 1 porque se identifica con las características computacionales de un sistema físico. Pero, puesto que uno de los requisitos de la investigación es el fisicalismo de los comportamientos, la siguiente cuestión que se ha abordado ha sido qué tipo de naturaleza física posee la fasa. Es decir, ¿la fasa se corresponde con una realidad microscópica o se trata de una propiedad que aparece en los sistemas macroscópicos como ocurre con la temperatura? Esta cuestión ha sido abordada en el capítulo 5. Las conclusiones de la discusión llevada a cabo es que la fasa tiene una base física, ya que las investigaciones que se están llevando a cabo en física fundamental apuntan a que para comprender el universo es necesario el uso de los conceptos de computación e información.

Una cuestión importante es que si los sistemas que poseen fasa son además unidades de evolución, la selección natural debería influir en los valores de fasa que presentan los sistemas biológicos. Así, en el capítulo 4 se presentaron diversas leyes que explicarían fenómenos de selección natural provocados por el tipo de fasa. La TGE propone que el valor de fasa  $\mathcal{F}$  es conformado por el modelo computacional, los datos que posee y la implementación física del modelo computacional. Por lo tanto, si el valor de fasa es determinado por fuerzas evolutivas en los sistemas biológicos se han planteado las siguientes preguntas: ¿Cómo representan información los sistemas nerviosos de los animales?, ¿qué proceso evolutivo ha determinado la representación de información en las redes neuronales biológicas? y ¿cuál es el origen evolutivo del modelo de red neuronal en el sistema nervioso? Las dos primeras preguntas fueron abordadas en el capítulo 6; aunque ninguna de ellas ha podido ser respondida porque existen muy pocos datos experimentales para poder abordar tanto el cómo se representan los estados en la actividad neuronal de los sistemas nerviosos como qué fuerzas evolutivas han influenciado el método que poseen los sistemas nerviosos de los animales para realizarlo. Aun así, la investigación ha llevado a proponer un nuevo mecanismo teórico para interpretar dinámicas cerebrales que independientemente de que sea usado o no en los sistemas nerviosos podría usarse en el diseño de sistemas conexionistas. La tercera pregunta se ha abordado en el capítulo 7. La investigación que se ha llevado a cabo ha concluido que la robustez computacional es una ventaja evolutiva que ha provocado la selección del modelo de red neuronal frente a los modelos computacionales que carecen de ella en los animales. Aun así, queda abierta la cuestión de si el modelo de red neuronal es el único biológicamente plausible con robustez computacional al daño o existe otra ventaja evolutiva que ha determinado la selección del modelo de red neuronal frente a otros modelos computacionales que pudieran tener también robustez computacional.

La última cuestión teórica que se ha abordado sobre la TGE ha sido si las limitaciones de las que se acusa al paradigma computacional de las ciencias cognitivas afectan a la TGE. Las ciencias cognitivas propusieron definir la mente como un proceso computacional para evitar el dualismo y emerger como una ciencia. La TGE evita el uso de conceptos del mentalismo para explicar el comportamiento, pero comparte con las ciencias cognitivas la noción de computación como elemento para explicar el

comportamiento. El paradigma computacional ha sido criticado por distintos filósofos y lógicos por haberse argumentado que tiene limitaciones en su capacidad de explicar la mente debido a argumentos relacionados con la importancia de la semántica. Así, la cuestión que se ha investigado es: ¿son los problemas semánticos una limitación real a la capacidad de explicar de la TGE? Los resultados de la investigación realizada sobre esa cuestión se han presentado en el capítulo 8. Una de las conclusiones ha sido que se ha ignorado la naturaleza computacional de las semánticas definidas por Tarski (las semánticas  $t$ ), y que estas se deben incluir en las descripciones computacionales tanto de la TGE como del paradigma computacional, ya que la inclusión de las semánticas  $t$  permiten mostrar que sistemas que son iguales en su descripción operacional, son computacionalmente diferentes. Esta conclusión es de vital importancia para la TGE, pero también para la ciencia cognitiva porque expande el paradigma computacional.

El hallazgo de la importancia de las semánticas  $t$  ha llevado a considerar la necesidad de que un marco matemático para la TGE debe ser capaz de manejar y expresar semánticas  $t$ . Así, se ha propuesto que la teoría cognitiva de condiciones de verdad (TCCV) es un marco matemático adecuado para describir valores de fase que causarían los comportamientos complejos que llevan a cabo seres humanos. Por ello, en el capítulo 9 se ha realizado una introducción a la TCCV. Mediante el marco de la TCCV se ha comenzado un programa de investigación que busca valores de fase para lograr robots multifuncionales mediante nociones topológicas (RMMNT). La investigación ha profundizado en la heurística de semántica topológica cualitativa. Una heurística que permite seleccionar acciones para establecer relaciones espaciales. Conforme a esa heurística se ha desarrollado una arquitectura para la navegación en espacios dinámicos y desconocidos abiertos que se ha denominado arquitectura de navegación cualitativa topológica (ANCT)1.0. El sistema de representación de relaciones espaciales propuesto inicialmente tiene limitaciones que impiden tomar decisiones sobre giros. Así, se ha desarrollado un sistema alternativo de coordenadas espaciales que sí permite tomar decisiones sobre giros. Además, se ha comenzado la investigación de métodos para la generación y procesamiento de lenguaje natural dentro del RMMNT en base a la TCCV. Toda la investigación referente al RMMNT ha sido presentada en el capítulo 10.

Todos los resultados obtenidos en esta tesis han sido obtenidos e interpretados en el marco de la TGE. Sin embargo, múltiples resultados de los obtenidos tienen interés por sí mismos. La importancia de las semánticas- $t$  para describir procesos computacionales es un resultado que es relevante para las ciencias cognitivas y la filosofía de la mente. La robustez computacional como ventaja evolutiva es un resultado importante para el campo que estudia la evolución del sistema nervioso. Las heurísticas cualitativas y la arquitectura de navegación cualitativa topológica son resultados relevantes para el campo de la robótica y la inteligencia artificial. Así, aunque la TGE ha servido de guía para dirigir la investigación, la validez de los resultados citados no depende de la validez de la TGE.

## 11.1. Conclusiones sobre la TGE

La TGE, a diferencia de teorías anteriores, parte de los fenómenos físicos observables y no de los fenómenos mentales que experimentan los seres humanos para explicar el comportamiento. Esta decisión viene provocada por el objetivo de formular una teoría fisicalista que unifique los comportamientos. Así, la TGE ni plantea ni busca explicar los estados mentales. La TGE propone la existencia de estados internos y procesos computacionales, pero la existencia de esos elementos se deriva de la asignación de una causa a los fenómenos que se busca explicar, no de la autoconsciencia del ser humano. Respecto al poder explicativo de la TGE, una cuestión que aborda la teoría es que los comportamientos intervienen en procesos evolutivos de extinción y supervivencia de organismos. La psicología y la IA son disciplinas que han buscado teorías para explicar, predecir y generar comportamientos. Sin embargo, no se han propuesto conexiones con los procesos evolutivos. Curiosamente la IA sí investiga cómo la selección natural puede ser usada para generar algoritmos que produzcan comportamientos complejos. Sin duda, esta diferencia hace que la TGE proponga una visión más completa de la naturaleza de lo que lo han hecho teorías anteriores de la psicología o la IA. Cabe mencionar también, que la TGE respecto a las teorías de psicología o de IA, más que una teoría que compite contra ellas es un marco en el que cada teoría de psicología o IA se interpretan como valores de fasa. Esto sin duda, es un hecho importante porque la TGE no significa desechar todo lo que se ha investigado en estas disciplinas sino reinterpretar su rol en un marco más amplio.

Otra cuestión importante es que existe una diferencia fundamental entre la TGE y el enfoque cognitivo para explicar los comportamientos. Esta es que, aunque el enfoque cognitivo determina que los procesos mentales son procesos computacionales, estos siempre llevan a cabo procesamiento de información; pero en cambio cuando la TGE determina que la causa de los comportamientos son procesos computacionales no se restringe a llevar a cabo procesamiento de información. Por lo tanto, en la TGE los procesos computacionales pueden ser o no con señales de entrada y no tienen por qué llevar a cabo procesamiento de información.

La IA actualmente carece de unos fundamentos científicos que la establezcan como una ciencia de la naturaleza. Tal vez alguien podría argumentar que una teoría que no usa el concepto de inteligencia no puede ser compatible con la IA. Sin embargo, situaciones similares se han dado ya en la ciencia. Un caso claro es el de la química orgánica, que tras la aparición de la teoría atómica moderna cambió su significado pasando a ser la de la química del carbono. Un campo científico debe ser refinado si esto permite una mejor comprensión. La TGE aporta unos fundamentos a la IA para que se ocupe de fenómenos naturales. Así, desde ese punto de vista la TGE merece seguir siendo investigada.

Algunas de las características más importantes de la TGE que se han observado

son las siguientes:

1. La TGE define un conjunto de fenómenos naturales, denominados exocomportamientos, que contiene los comportamientos de sistemas biológicos, no biológicos autónomos, y los fenómenos de plasticidad fenotípica.
2. La TGE determina que existe sólo una propiedad como causa de los exocomportamientos, denominada fasa.
3. La TGE determina que existen diferente tipos de valores fasa, lo que causa diferentes tipos de exocomportamientos.
4. La TGE determina que cada valor de fasa es un proceso computacional.
5. La TGE se trata de una teoría fiscalista.
6. La TGE conecta la física fundamental con los comportamientos.
7. La TGE no sólo se ocupa de predecir y generar exocomportamientos, sino que explica y predice fenómenos evolutivos.
8. La TGE sirve como marco general para teorías o mecanismos de generación de comportamientos.

Tras la investigación realizada en esta tesis, se puede comparar la situación de la explicación de los comportamientos con la de la física a finales del siglo XIX explicando las ondas electromagnéticas. Después de la explicación del electromagnetismo de Maxwell, se postuló la existencia del éter. El éter se propuso como la sustancia material que permitiría a las ondas electromagnéticas propagarse, ya que parecía difícilmente concebible que una onda se propagase en el vacío sin ningún medio material que hiciera de soporte. En el caso de los comportamientos ha sucedido una situación similar a la de la física con el éter. Los seres humanos hemos considerado que los comportamientos complejos de los seres humanos necesitan de un elemento especial que solo poseen los seres humanos, al que se ha denominado inteligencia. Sin embargo, todas las investigaciones nos conducen a considerar que ese elemento que separa a los humanos del resto de los organismos no existe. El principio de la navaja de Ockham nos dice que no debemos de multiplicar las entidades en una explicación cuando se puede dar una explicación con menos entidades sin que haya diferencia entre una explicación u otra. Así, la TGE parece ser una teoría que en base al principio de la navaja de Ockham resulta ser una mejor explicación de los comportamientos que las teorías de la inteligencia.

## 11.2. Conclusiones sobre la TCCV y el programa RMMNT

La TCCV se ha usado como base para guiar el programa RMMNT. Esto hace que las conclusiones sobre la TCCV que surgen de esta investigación estén ligadas a los resultados del programa RMMNT. Dentro del programa RMMNT, se ha logrado desarrollar una arquitectura que permite la navegación en tiempo real en entornos desconocidos y dinámicos en base a establecer relaciones espaciales cualitativas. Además, se ha desarrollado un sistema de coordenadas cualitativas esféricas que permite tomar decisiones de giro. Tras estos mecanismos se ha comenzado una nueva fase para el desarrollo de métodos de generación y procesamiento de lenguaje natural. Por lo tanto, el programa RMMNT es sin duda una línea de investigación en la que continuar trabajando, ya que se han desarrollado nuevos algoritmos y métodos que pueden ser usados en robótica. Esto hace concluir que es factible la hipótesis que sostiene la TCCV de que las condiciones de verdad son elementos fundamentales en todas las características cognitivas que el cerebro humano posee. Existe la posibilidad de que el cerebro humano usara métodos o mecanismos no basados en condiciones de verdad. Pero, desde el punto de vista de la TGE, un exocomportamiento podría ser generado por diversos valores de fasa. Por lo tanto, la TCCV es cuando menos un marco teórico que permite crear arquitecturas cognitivas que dotan de capacidades cognitivas útiles para integrar robots en la vida cotidiana de los seres humanos.

## 11.3. Líneas futuras de investigación

Las líneas futuras que emergen tras la realización de las investigaciones presentadas en esta tesis están situadas tanto en el ámbito de la ciencia básica como de la aplicada. En el caso de la TGE, los resultados obtenidos han mostrado que pese a ser una teoría que apenas acaba de nacer ya proporciona una nueva visión de los comportamientos como fenómenos naturales. Así, continuar la investigación de la TGE es sin duda una línea totalmente justificada. Existen varias posibilidades distintas para continuar la investigación de la TGE. Una de ellas es estudiar las leyes que relacionan fasa con fenómenos evolutivos de extinción y supervivencia. Caben dos posibilidades. Una es buscar algún tipo de unificación de las leyes que produzca una ley que contemple las leyes formuladas hasta ahora como casos particulares. Otra línea es investigar la existencia de nuevas leyes que relacionen fasa con otros fenómenos evolutivos. Otra cuestión importante de esta línea de investigación sería desarrollar experimentos con los que poner a prueba las leyes y que ayuden también a comprender más sus significados y consecuencias en la naturaleza.

Otra importante línea de investigación básica que emerge de esta tesis es la inves-

tigación de la robustez computacional como una ventaja evolutiva. La investigación realizada se ha basado en razonamientos lógicos y experimentos mentales. Así, la investigación se puede continuar con simulaciones evolutivas para comprobar si estas son coherentes con los razonamientos y respaldan los resultados ya obtenidos.

Otra línea de investigación es la TCCV. En este caso se trata de una línea que se encuentra entre la investigación básica y aplicada, ya que, aunque la TCCV es un marco matemático, este sirve como base para desarrollar algoritmos y arquitecturas para el programa RMMNT. Cuestiones como ¿existen conceptos que están en la representación de la realidad que tienen los seres humanos que no están en la estructura PPFMM?, ¿habría que añadir más lenguajes a la jerarquía? y ¿necesitan más símbolos los lenguajes de la jerarquía?, son temas sobre los que se puede investigar.

Finalmente, la última línea de investigación tras esta tesis es la que surge de la investigación en el programa RMMNT. Los resultados obtenidos han mostrado que es posible la navegación cualitativa basada en relaciones espaciales y ha producido el comienzo de una segunda fase para el procesamiento y generación de lenguaje natural. Por lo tanto, se han sentado las bases para investigar y desarrollar nuevos métodos cualitativos. Además, los algoritmos desarrollados de la primera fase del programa RMMNT solo se han probado en entornos virtuales. Luego, los resultados habilitan el objetivo de investigar la implantación de los algoritmos y arquitecturas del RMMNT en robots físicos que se integren en la vida diaria de los seres humanos.



# Capítulo 12

## Conclusions

In this doctoral thesis, research to formulate and develop a theory to unify the behaviors of biological and nonbiological autonomous systems has been presented. A general review of the proposals explaining these behaviors was provided at the beginning of this thesis. This review, presented in chapter 2, concluded that the concepts of intelligence and cognition cannot properly allow for the unification of these behaviors because of the incompatibility of those concepts with the behaviors discovered in nature. Additionally, both concepts have been in a state of ambiguity for a long time, making it reasonable to investigate alternative ideas for explaining the behaviors. An additional review was provided in this chapter about the effects of the behaviors on the evolutionary process 3. This second review showed that, instead of being isolated phenomena, the behaviors drive the direction of evolution, and it was concluded that a theory to unify the behaviors should be approached as if the behaviors determine the direction of evolution. Thus, given that the proposed theories to explain the behaviors focus only on the behavior itself, and that there is a need to investigate alternative concepts of intelligence, the current research has focused on the General Theory of Exobehaviors (GTE) to unify the behaviors. The GTE is a recent proposal that was initially used to obtain a more precise formulation than what existed. The new formulation, presented in chapter 4, allows for differentiating phenomena that previous formulations could not. The GTE defines a combination of physical phenomena, called exobehaviors. In addition to the behaviors of biological and nonbiological autonomous systems, the group of exobehaviors also includes the structural development phenomena of biological systems. The GTE proposes a property called *fass* causes exobehaviors. The formulation of *fass* meets the requirements of physicalism presented in chapter 1 because it identifies the computational characteristics of a physical system. However, because one of the requirements of the investigation is physicalism, the question of what type of physical nature *fass* possesses has been raised. Namely, does *fass* correspond to a microscopic reality or does it deal with a property that appears in macroscopic systems, such as occurs with temperature? This question was addressed in chapter 5,

and that discussion concluded that fass has a physical base since the research that is underway in fundamental physics points to the use of the concepts of computation and information science as necessary for understanding the universe.

An important question to consider is that if systems with fass are also units of evolution, should natural selection affect the fass values that biological systems present? In chapter 4, various laws explaining aspects of natural selection caused by fass were discussed. The GTE proposes that fass value  $\mathcal{F}$  is defined by the computational model, the data it possesses, and the physical implementation of the computational model. Therefore, if the fass value is determined by evolutive force in biological systems, the following questions arise: How do nervous systems represent information about animals? What evolutionary process determines the representation of information in biological neuronal networks? What is the evolutionary origin of the neuronal network model in the nervous system? Although the first two questions were addressed in chapter 6, neither of them could be answered because too little experimental data exists to address both how the states of neuronal activity in the nervous system are represented and which evolutionary forces have influenced the method nervous systems of animals possess to do this. Nevertheless, the research led to a proposal of a new theoretical mechanism for interpreting cerebral dynamics that, regardless of whether it is used for nervous systems, could be used in the design of connectionist systems. The third question was addressed in chapter 7. The investigation concluded that computational robustness is an evolutionary advantage that leads to the selection of the neuronal network model in animals against computational models that lack this advantage. Despite this, the question remains whether the neuronal network model is the only biologically plausible network with computational robustness to damage or if another evolutionary advantage with computational robustness exists that selects the neuronal network model above other computational models. The last theoretical question addressed by the GTE was if the supposed limitations of the cognitive science computational paradigm affect the GTE. The field of cognitive science proposes to define the mind as a computational process to avoid dualism and to emerge as a science. The GTE avoids the use of mentalism to the explain behaviors, and instead, like cognitive science, uses the idea of computation to explain behavior. The computational paradigm has been criticized by some philosophers and logicians for having limitations in its capacity to explain the mind because of arguments related to the importance of semantics. Thus, the following question remains: Are semantic problems a true limitation in explaining the GTE? The results of the investigation regarding this question were presented in chapter 8. One of the conclusions was that the computational nature of semantics defined by Tarski (*t*-semantics) was disregarded and that it should have been included in the computational descriptions in both the GTE and the computational paradigm, since its inclusion allows for showing that systems equal in operational description are computationally different. This conclusion is vitally important for the GTE, but also for cognitive science because it expands the computational paradigm. The discovery of

the importance of t-semantics led to the requirement for a mathematical framework for the GTE that is capable of managing and expressing t-semantics. It has been proposed that the Cognitive Theory of True Conditions (CTTC) is an appropriate mathematical model to describe the fass values that caused complex behaviors to come about in humans. For this reason, an introduction to the CTTC was provided in chapter 9. Through the framework of the CTTC, a research program was created to search for fass values to develop multifunctional robots through topological notions (MROTN). The research has examined the heuristic of qualitative topological semantics in depth. This heuristic allows for classifying actions to establish spatial relations. In accordance with this heuristic, an architecture to navigate dynamic and unknown open spaces has been developed, which is called Architecture of Qualitative Topological Navigation (AQTN 1.0). Initially, the proposed system of representation of spatial relations faced limitations that impeded making decisions about turns. Therefore, an alternative system of spatial coordinates was developed to allow for making decisions about turns. Additionally, research based on the CTTC regarding methods for natural language generation and processing began within the MROTN. Research related to the MROTN was discussed in chapter 10.

The results obtained in this thesis were derived and interpreted under the framework of the GTE. However, many of the results are interesting in and of themselves. The importance of t-semantics in describing computational processes is relevant to cognitive science and philosophy of mind. Computational robustness as an evolutionary advantage is important within the field of nervous system evolution. The qualitative heuristics and qualitative topological navigation architecture are findings relevant to the field of robotics and artificial intelligence. Thus, even though the GTE has served as a guide for directing the research, the validity of the cited results does not depend on the validity of the GTE.

## 12.1. Conclusions about the GTE

The GTE, unlike previous theories, starts from observable physical phenomena to explain behavior rather than mental phenomena that we experience as human beings. This decision was made to formulate a physicalist theory that unifies the behaviors. The GTE neither considers nor tries to explain mental states. It proposes the existence of both internal states and computational processes, but the existence of these elements is derived from assigning a cause to the phenomena being sought to explain and not from the self-consciousness of humans. Regarding the explanatory power of the GTE, the theory addresses whether the behaviors intervene in the evolutionary processes of extinction and survival. Psychology and AI have tried to explain, predict, and generate behaviors, but these fields have not proposed connections with evolutionary processes. Interestingly, AI does investigate how natural selection can be used to generate algo-

rhythms that produce complex behaviors. Without question, this difference causes the GTE to propose a more complete view of nature than previous theories within psychology and AI. It should also be mentioned that the GTE, rather than competing with theories within psychology or AI, is a framework through which these theories can be interpreted as fass values. This, without a doubt, is important because the GTE does not dismiss what these other disciplines have investigated but instead reinterprets their role in an amplified framework. Another important question is whether a fundamental difference exists between the GTE and the cognitive approach for explaining behaviors. Although the cognitive approach concludes that mental processes are computational processes, they always carry out information processing; however, when the GTE determines that computational processes cause behaviors, it is not limited to information processing. Therefore, computational processes in the GTE can occur with or without input signals, and they do not have to carry out information processing. Currently, AI lacks a scientific foundation that establishes it as a natural science. Perhaps one could argue that a theory that does not use the concept of intelligence cannot be compatible AI. Nevertheless, similar situations have already occurred in science. A clear case is that of organic chemistry, which changed its significance to the chemistry of carbon after the appearance of modern atomic theory. A scientific field should be refined if it allows for a better understanding. The GTE provides a foundation for AI to deal with natural phenomena, and from this point of view, the GTE deserves further investigation. Some of the most important characteristics of the GTE are the following:

1. The GTE defines a set of natural phenomena, called exobehaviors, that includes the behaviors of biological systems, nonbiological autonomous systems, and the phenomena of phenotypic plasticity.
2. The GTE determines that only one property, called fass, exists as the cause of exobehaviors.
3. The GTE determines that different types of fass values exist, which cause different types of exobehaviors.
4. The GTE determines that each fass value is a computational process.
5. The GTE is a physicalist theory.
6. The GTE connects fundamental physics with behaviors.
7. The GTE not only deals with predicting and generating exobehaviors but also explaining and predicting evolutionary phenomena.
8. The GTE serves as a general framework for behavior-generating theories or mechanisms.

Following the research performed in this thesis, we can compare the situation of explaining behaviors with that explaining electromagnetic waves in physics at the end of the 19th century. After Maxwell's explanation of electromagnetism, the existence of ether was postulated. Ether was proposed as a material substance that could allow electromagnetic waves to propagate, as it seemed inconceivable that a wave could be propagated in a vacuum without any material substance to support it. In the case of behaviors, a similar situation occurred to that of ether in physics. Human beings have considered complex human behavior, which they call intelligence, as a unique and necessary element that only they possess. However, research leads us to believe that an element that separates humans from the rest of the organisms does not exist. Ockham's Razor tells us that entities should not be multiplied when an explanation can be given with fewer entities without there being a difference between the two explanations. Thus, the GTE, which is a theory based on Ockham's Razor, appears to better explain behaviors than theories of intelligence.

## 12.2. Conclusions about the TCCV and the MROTN program

The MROTN has used the CTTC as a base to guide the program. Therefore, conclusions about the CTTC that arise from the investigation are linked to the results of MROTN. Within the MROTN program, an architecture has been developed based on establishing qualitative spatial relationships that allows for navigation in unknown and dynamic environments in real time. Additionally, a spherical qualitative coordinate system has been developed that permits decision making about turns. After the development of these mechanisms, a new phase began for generating and processing natural language. Therefore, the MROTN program is a line of research in which to keep working, since new algorithms and methods have been developed that can be used in robotics. This leads to the conclusion that the hypothesis that maintains the true conditions of the CTTC are fundamental elements in all cognitive characteristics possessed by the human brain is feasible. The possibility exists that the human brain uses methods or mechanism not based on true conditions. From the point of view of the GTE, an exobehavior could be generated by various fass values; however, the CTTC is at least theoretical framework that allows for the creation of cognitive architectures that provide useful cognitive abilities for integrating robots into everyday human life.

## 12.3. Future Research Directions

Future research directions that emerge after the presentation of this thesis lie in the realms of both basic and applied sciences. Regarding the GTE, the results obtained

have shown that, despite being a newly created theory, it already provides a new view of behaviors as natural phenomena. Therefore, continuing to research the GTE is surely a justified research direction. Several different possibilities exist for continuing research on the GTE. One of these is to study the laws relating fitness to the evolutionary phenomena of extinction and survival. There are two possibilities. One is to search for a type of unification of laws to produce a law that considers laws formulated until now to be particular cases. The other is to investigate the existence of new laws that relate fitness values to other evolutionary phenomena. Another important issue in this line of research is to develop experiments to test and better understand the laws and their meanings and consequences in nature. Another important area of basic research that has emerged from this thesis is the investigation of computational robustness as an evolutionary advantage. The research performed was based in logical reasoning and mental experiments. Therefore, the research can continue with evolutionary simulations to test if it is consistent with the reasoning and supports the results already obtained. Another area of research to pursue is the CTTC, which is a line between basic and applied research. Although it is a mathematical framework, it serves as a base to develop algorithms and architectures for the MROTN program. The following questions are themes that should be further investigated: Are there concepts that represent the reality that human beings experience that are not in the MMPPF structures? Do the languages of the hierarchy need more symbols? Finally, the last line of research to further pursue is that which has arisen from the research of the MROTN program. The results obtained have shown that qualitative navigation based on spatial relations is possible, and a second phase of research regarding processing and generating natural language has begun. Therefore, the foundation has already been laid for the research and development of new quantitative methods. In addition, the algorithms developed in the first phase of the MROTN program have only been tested in virtual environments. Subsequently, these results enable further investigation into the implementation of algorithms and MROTN architectures in physical robots that are integrated into the daily life of human beings.

# Apéndice A

## Axiomas de la TCCV

Este apéndice tiene las definiciones de las estructuras PPFMM, las estructuras de perspectiva temporal, las estructuras estado y sus axiomas. Una estructura PPFMM es construida a partir de perspectivas temporales, y una estructura de perspectiva temporal es construida usando estructuras estado. Así, primero se define una estructura PPFMM y después se definen el resto descendiendo según la relación citada. Una vez realizadas las definiciones se dan los axiomas de la estructura PPFMM.

Una estructura PPFMM se define formalmente como la tupla  $\langle \mathcal{U}, T, E, I, \mathcal{D}^p, \rho_{PPFMM}, \langle d_j^p \rangle, L \& \textcircled{\mathbf{i}} \rangle$

cuyos dominios son:

$\mathcal{U} = \{\mathcal{P}_1, \dots, \mathcal{P}_m\}$  is the set of temporal perspectives

$T = \{1, \dots, m\}$  es el conjunto tiempo

$E = \{e_1, \dots, e_s\}$  es el conjunto de estados

$I = \bigcup_j I_j$  y  $I_j = \theta_j^0(o_1) \times \dots \times \theta_j^n(o_1) \times \dots \times \theta_j^0(o_z) \times \dots \times \theta_j^n(o_z) \quad j \in E$

$\mathcal{D}^p = \bigcup_j D_j^p$  y  $D_j^p = \{S^{o_i, p} o_u : S^{o_i, p} o_u \in e_j\}$  es el conjunto de dependencias

, sus funciones son

$$\rho_{PPFMM} : T \longrightarrow \mathcal{U}$$

$$d_j^p : \mathcal{P}(\mathcal{P}(H \times V_0) \times \mathcal{P}(H \times V_p) \times \mathcal{V}) \times \mathcal{P}(A^p \times \mathcal{V}) \longrightarrow \mathcal{D}^p \quad j \in E$$

$L = \{l^0, \dots, l^n\}$  es el conjunto de leyes del entorno.

$$l^p : \mathcal{P}(\mathcal{P}(H \times V_p) \times \mathcal{V}) \times \mathcal{P}(A^p \times \mathcal{V}) \times \mathcal{D}^p \longrightarrow \mathcal{P}(\mathcal{P}(H \times V_p))$$

$$\& : E \times I \longrightarrow E$$

$$\textcircled{i} : T \longrightarrow I$$

La definición de la estructura de perspectiva temporal,  $\mathcal{P}_t$ , is

$$\mathcal{P}_t = \langle M_{\mathcal{P}_t}, T_{\mathcal{P}_t}, E_{\mathcal{P}_t}, I_{\mathcal{P}_t}, \mathcal{D}_{\mathcal{P}_t}^p, \langle d_j^p \rangle, L_{\mathcal{P}_t}, SL_{\mathcal{P}_t}, \&_{\mathcal{P}_t}, \textcircled{i}_{\mathcal{P}_t}, Succ_{\mathcal{P}_t}, \prec_{\mathcal{P}_t} \rangle$$

cuyos dominios son:

$$M_{\mathcal{P}_t} = \{(t', \diamond, \square, e_x)\} \quad \diamond \in \mathfrak{C} = \{h, \varepsilon\} \quad \square \in \mathfrak{S} = \{\downarrow, |, ||, |\downarrow\} \quad t' \in T_{\mathcal{P}_t} \quad e_x \in E_{\mathcal{P}_t}$$

- If  $t > t'$  then  $\square = \downarrow$
- If  $t = t'$  then  $\square = ||$
- If  $t < t'$  then  $\square = |\downarrow$

$T_{\mathcal{P}_t} = \{1, \dots, m\}$  es el conjunto tiempo

$E_{\mathcal{P}_t} = \{e_1, \dots, e_s\}$  es el conjunto de estados

$$I_{\mathcal{P}_t} = \bigcup_j I_j \text{ y } I_j = \theta_j^0(o_1) \times \dots \times \theta_j^n(o_1) \times \dots \times \theta_j^0(o_z) \times \dots \times \theta_j^n(o_z) \quad j \in E$$

, las funciones son:

$$\mathcal{D}_{\mathcal{P}_t}^p = \bigcup_j D_j^p \text{ and } D_j^p = \{S_j^{o_i, p} o_u : o_u \in e_j\}$$

$L_{\mathcal{P}_t} = \{l^0, \dots, l^n\}$  es el conjunto de leyes del entorno de  $\mathcal{P}_t$ .

$$l^p : \mathcal{P}(\mathcal{P}(H \times V_p) \times \mathcal{V}) \times \mathcal{P}(A^p \times \mathcal{V}) \times \mathcal{D}^p \longrightarrow \mathcal{P}(\mathcal{P}(H \times V_p))$$

$$SL_{\mathcal{P}_t} = \{sl^1, \dots, sl^r\}$$

$$sl^k : SRA \times \mathcal{V} \times E_{\mathcal{P}_t} \times I_{\mathcal{P}_t} \longrightarrow SRA$$

$$\begin{aligned} \&_{\mathcal{P}_t} : E \times I &\longrightarrow E \\ \textcircled{i}_{\mathcal{P}_t} : T &\longrightarrow I \\ Succ_{\mathcal{P}_t} : (T \times \mathfrak{C} \times \mathfrak{S} \times E) \times I &\longrightarrow (T \times \mathfrak{C} \times \mathfrak{S} \times E) \end{aligned}$$

, y la relación  $\prec_{\mathcal{P}_t}$  es

$$\prec_{\mathcal{P}_t} = \{ \langle (t, \diamond, \square, e_x), (t', \diamond, \square, e_x) \rangle : t < t' \text{ and } (t, \diamond, \square, e_x), (t', \diamond, \square, e_x) \in M_{\mathcal{P}_t} \}$$

La definición de una estructura de estado,  $e_j$ , es

$$e_j = \langle \langle U_i \rangle \langle fl \rangle \langle R_k \rangle \rangle$$

cuyos dominios son:

$$\langle U_i \rangle = \langle \mathcal{V}, \mathcal{P}(H), \mathcal{P}(H \times V_0), \dots, \mathcal{P}(H \times V_n), A_1^p, \dots, A_z^p, SRA \rangle$$

$$\begin{aligned} \mathcal{V} &= \{o_1, \dots, o_z\} \\ H &= \{h_1, \dots, h_{z'}\} \quad z \leq z' \\ V_p &= \{\emptyset\} \cup \{(w_1, \dots, w_c) : w_1 \in W_{p,1}, \dots, w_c \in W_{p,c}\} \\ A_i^p &= \{a_{i,1}^p, \dots, a_{i,k}^p\} \quad i \in \mathcal{V} \text{ and } a_{i,k}^p = (a_{i,k}^p.in, a_{i,k}^p.ext) \\ a_{i,k}^p.in &: DH_{p,k} \times DV_{p,k} \longrightarrow CV \quad DH_{p,k} \subseteq H, DV_{p,k}, CV_{p,k} \subseteq V_p \\ a_{i,k}^p.ext &: H_{p,k} \times DV_{p,k} \longrightarrow CV \quad H_{p,k} \subseteq H, DV_{p,k}, CV_{p,k} \subseteq V_p \\ SRA &= \{ \langle (s_1, c_1), \dots, (s_r, c_r) \rangle, \dots, \langle (c_1, c_1), \dots, (c_r, c_r) \rangle \} \cup \{\emptyset\} \end{aligned}$$

, las funciones son

$$\langle fl \rangle = \langle ES_j, g_j^{*0}, \dots, g_j^{*n}, g_j^0, \dots, g_j^n, \theta_0^p, \dots, \theta_n^p, \odot \rangle$$

$$ES_j : \mathcal{V} \longrightarrow \mathcal{P}(H)$$

$$g^{*p} : \mathcal{V} \longrightarrow \mathcal{P}(H \times V_p)$$

$$g^p : H \longrightarrow V_p$$

$$\theta^p : \mathcal{V} \longrightarrow \cup_i \mathcal{P}(A_i^p)$$

$$\odot : \mathcal{V} \longrightarrow SRA$$

, y las relaciones son:

$$\begin{aligned} \langle R_k \rangle &= \langle S_j^{o_1,0}, \dots, S_j^{o_1,n}, \dots, S_j^{o_z,0}, \dots, S_j^{o_z,n} \rangle \\ S_j^{o_i,p} &\subseteq \mathcal{V} \end{aligned}$$

Los siguiente axiomas definen una estructura PPFMM:

■ **Axioma 1°**

$$\forall e_j \in E_{\mathcal{P}_t} \quad (g_j^p(h) \neq \emptyset \wedge p > 0 \Leftrightarrow \{h\} \subset ES_j(o_i))$$

El primer axioma determina que un elemento esencia tiene asignado un valor de alguna propiedad si y solo si esta asignado a un objeto.

■ **Axioma 2°**

$$\forall e_j \in E_{\mathcal{P}_t} \quad (p > 0) \wedge (g_j^p(h) \neq \emptyset \Leftrightarrow g_j^0(h) \neq \emptyset)$$

El segundo axioma determina que un lugar del espacio es únicamente asignado a un objeto si y solo si tiene asignado algún valor de otra propiedad.

■ **Axioma 3°**

$$\forall e_j \in E_{\mathcal{P}_t} \quad (g_j^p(h) = x \Leftrightarrow \exists o_i(h, x) \in g_j^{*p}(o_i))$$

El tercer axioma determina la relación entre  $g_j^p$  y  $g_j^{*p}$ .

■ **Axioma 4°**

$$\forall e_j \in E_{\mathcal{P}_t} \quad (a_{i,k}^p \in A_i^p \Rightarrow DH_{p,k} = \bigcup_j ES_j(o_i))$$

El cuarto axioma asegura que una acción asegura independientemente of the assignation done by  $ES_j$ .

■ **Axioma 5°**

$$\forall e_j e_{j'} \in E_{\mathcal{P}_t} \quad (h \in ES_j(o_i) \wedge h \in ES_{j'}(o_u) \Leftrightarrow u = i)$$

El quinto axioma determina que un elemento esencia solo es asignado a un objeto.

■ **Axioma 6°**

$$Succ_{\mathcal{P}_t}(t, e_j) = e_{j'} \Leftrightarrow d_j^p(\langle (g_j^{*0}(o_i), o_i) \rangle_i, \langle (g_j^{*p}(o_i), o_i) \rangle_i, \langle (a_{i,k}^{*p}, o_1) \rangle_i) = D_{j'}^p$$

El sexto axioma determina que las dependencias son coherentes con cambios de  $e_j$  a  $e_{j'}$ .

■ **Axioma 7°**

$$Succ_{\mathcal{P}_t}(t, e_j) = e_{j'} \Leftrightarrow \exists \langle a_{i,k}^p, o_i \rangle_i \quad \forall p \langle (g_j^{*p}(o_i), o_i) \rangle_i = l^p(\langle (g_j^{*p}(o_i), o_i) \rangle_i, \langle a_{i,k}^p, o_i \rangle_i, D_j^p)$$

El axioma septimo determina que si el estado  $e_{j'}$  sucede a  $e_j$  es porque los objetos pueden producir cambios que generan  $e_{j'}$  a partir de  $e_j$ .

■ **Axioma 8°**

$$\forall e_j \in E_{\mathcal{P}_t} \quad \theta_j^p(o_i) \in \mathcal{P}(A_i^p)$$

El axioma octavo determina que  $\theta_j^p$  solo asigna acciones a un objeto si ellos modifican que objeto.

■ **Axioma 9°**

$$\begin{aligned} \forall t \in T \quad T_{\mathcal{P}_t} &= T \\ \forall t \in T \quad E_{\mathcal{P}_t} &= E \\ \forall t \in T \quad I_{\mathcal{P}_t} &= I \\ \forall t \in T \quad L_{\mathcal{P}_t} &= L \\ \forall t \in T \quad \&_{\mathcal{P}_t} &= \& \\ \forall t \in T \quad \mathbf{i}_{\mathcal{P}_t} &= \mathbf{i} \\ \forall t \in T \quad \mathcal{D}_{\mathcal{P}_t}^p &= \mathcal{D}^p \\ \forall t \in T \quad \langle d_j^p \rangle_{\mathcal{P}_t} &= \langle d_j^p \rangle \\ \forall \mathcal{P}_t, \mathcal{P}_{t'} \quad |SL_{\mathcal{P}_t}| &= |SL_{\mathcal{P}_{t'}}| \\ \forall \mathcal{P}_t, \mathcal{P}_{t'} \quad sl_{\mathcal{P}_t}^k &= sl_{\mathcal{P}_{t'}}^k \end{aligned}$$

El axioma noveno determina que los dominios no cambian de una perspectiva temporal a otra.

■ **Axioma 10°**

$$\forall \mathcal{P}_t \quad Succ_{\mathcal{P}_t}((t', \diamond, \square, e_x), i) = \begin{cases} (t' + 1, \varepsilon, \downarrow, e_y), & \text{for } t' + 1 < t \text{ and } \mathbf{i}(t') = i_j \\ (t' + 1, h, \downarrow, e_y), & \text{for } t' + 1 < t \text{ and } \mathbf{i}(t') \neq i_j \\ (t' + 1, \varepsilon, ||, e_y), & \text{for } t' + 1 = t \text{ and } \mathbf{i}(t') = i_j \\ (t' + 1, h, ||, e_y), & \text{for } t' + 1 = t \text{ and } \mathbf{i}(t') \neq i_j \\ (t + 1, \varepsilon, | \downarrow, e_y), & \text{for } t' \geq t \text{ and } \diamond = \varepsilon \\ (t + 1, h, | \downarrow, e_y), & \text{for } t' \geq t \text{ and } \diamond = h \end{cases}$$

El axioma décimo determina en que periodo de tiempo está situado un momento de tiempo y su condición de realidad.

■ **Axioma 11**

$$\forall e_j, e_{j'} \in E_{\mathcal{P}_t} \quad Succ_{\mathcal{P}_t}((t, \diamond, \square, e_j), i) = (t, \diamond, \square, e_{j'}) \rightarrow \odot_{j'}(o_i) = sl^k(\odot_j(o_i), o_i, e_j, i)$$

El axioma onceavo determina que los cambios de sensaciones de recompensa y castigo es coherente con las acciones que el objeto lleva a cabo.



# Apéndice B

## Alfabetos de metainformación

Este apéndice define los alfabetos de metainformación  $\Sigma_1$  y  $\Sigma_2$ , y muestra cómo se construyen las fórmulas del lenguaje  $L_{PPFMM}^*$  a partir de una fórmula de  $L_{PPFMM}^*$ . Esto permitirá hacerse una idea más clara del tipo de abstracción que ocurre cuando se asciende de lenguaje en la jerarquía.

### Alfabeto de metainformación $\Sigma_1$

El alfabeto de metainformación  $\Sigma_1$  contiene dos elementos que se denotan por 0 y 1. Es decir,

$$\Sigma_1 = \{0, 1\}$$

Los elementos de los alfabetos de metainformación son relaciones. Así, su definición es la siguiente:

- $\forall o_i, p, g_j^{*p} \quad 0\langle o_i, p, g_j^{*p}(o_i) \rangle \longleftrightarrow g_j^{*p}(o_i) = \emptyset$
- $\forall o_i, p, g_j^{*p} \quad 1\langle o_i, p, g_j^{*p}(o_i) \rangle \longleftrightarrow g_j^{*p}(o_i) \neq \emptyset$

Se llama estado momentáneo a cada elemento del conjunto  $\Sigma_1$ . Mediante  $\Sigma_1$  se da un estado cualitativo de un objeto en un momento del tiempo de la Propiedad-p, informando de si un objeto tiene una determina cualidad, o no la tiene, sin importar el valor que tenga. A continuación, se define un funtor del lenguaje  $L_{PPFMM}^*$  que asocia a cada objeto en un momento del tiempo y respecto a una cualidad un elemento del alfabeto de metainformación.

Se llamará *estado momentáneo del componente-p-ésimo* al funtor  $em_p$  que se define de la siguiente manera:

- $\forall o_i, p \quad em_p(o_i, \psi) = 0 \longleftrightarrow g_j^{*0}(o_i) = \emptyset$

$$\blacksquare \forall o_i, p \quad em_p(o_i, \psi) = 0 \longleftrightarrow g_j^{*0}(o_i) \neq \emptyset$$

donde  $\psi \equiv [\square |o_i| g_j^{*0}(o_i)|, \dots, |g_j^{*n}(o_i)|]$  y  $\square$  el posicionador temporal correspondiente. Usando el funtor  $em_p$  se puede construir la fórmula siguiente:

$$\psi^* \equiv [\square |o_i| em_0(o_i, \psi)|, \dots, |em_n(o_i, \psi)|]$$

Así,  $\psi^*$ , que pertenece a  $L_{PPFMM}^*$ , se ha construido a partir de  $\psi$ .

## Alfabeto de meta información $\Sigma_2$

El alfabeto de meta información  $\Sigma_2$  contiene tres elementos que se denotan por  $\wr$ ,  $\beta_1$  y  $\beta_2$ . Es decir,

$$\Sigma_2 = \{\wr, \beta_1, \beta_2\}$$

Como se ha dicho, los elementos de los alfabetos de meta información son relaciones. Así, su definición es la siguiente:

- $\forall o_i, p, g_j^{*p}, g_{j'}^{*p} \quad \beta_1 \langle o_i, p, g_j^{*p}(o_i), g_{j'}^{*p} \rangle \longleftrightarrow g_j^{*p}(o_i) = g_{j'}^{*p}(o_i) \wedge \phi$
- $\forall o_i, p, g_j^{*p}, g_{j'}^{*p} \quad \beta_2 \langle o_i, p, g_j^{*p}(o_i), g_{j'}^{*p} \rangle \longleftrightarrow g_j^{*p}(o_i) \neq g_{j'}^{*p}(o_i) \wedge \phi$
- $\forall o_i, p, g_j^{*p}, g_{j'}^{*p} \quad \wr \langle o_i, p, g_j^{*p}(o_i), g_{j'}^{*p} \rangle \longleftrightarrow em_p(o_i, \psi) = 0 \vee em_p(o_i, \varphi) = 0$

donde  $\phi \equiv em_p(o_i, \psi) \neq 0 \wedge em_p(o_i, \varphi) \neq 0$

y

$$\psi \equiv [\square |o_i| g_j^{*0}(o_i)|, \dots, |g_j^{*n}(o_i)|]$$

$$\varphi \equiv [\square |o_i| g_{j'}^{*0}(o_i)|, \dots, |g_{j'}^{*n}(o_i)|]$$

Llamaremos *estado temporal de la propiedad-p-ésima* al funtor  $etp^p$  que se define del siguiente modo:

- $\forall o_i, \varphi \quad etp^p(o_i, p, \varphi) = \beta_1 \longleftrightarrow g_j^{*p}(o_i) = g_{j'}^{*p}(o_i) \wedge \phi$
- $\forall o_i, p, \varphi \quad etp^p(o_i, p, \varphi) = \beta_2 \longleftrightarrow g_j^{*p}(o_i) \neq g_{j'}^{*p}(o_i) \wedge \phi$
- $\forall o_i, p, \varphi \quad etp^p(o_i, p, \varphi) = \wr \longleftrightarrow em_p(o_i, \psi) = 0 \vee em_p(o_i, \psi') = 0$

donde

$$\phi \equiv em_p(o_i, \psi) \neq 0 \wedge em_p(o_i, \varphi) \neq 0$$

y

$$\varphi \equiv \psi \blacktriangleright \psi'$$

y

$$\psi \equiv [\square |o_i| g_j^{*0}(o_i), \dots, |g_j^{*n}(o_i)|]$$

$$\psi' \equiv [\square |o_i| g_{j'}^{*0}(o_i), \dots, |g_{j'}^{*n}(o_i)|]$$

Usando el funtor  $etp^p$  se puede construir la fórmula siguiente:

$$\varphi^* \equiv [\square |o_i| etp^0(o_i, \varphi), \dots, |etp^n(o_i, \varphi)|]$$

Así,  $\varphi^*$ , que pertenece a  $L_{PPFMM}^*$ , se ha construido a partir de  $\varphi$ .



# Apéndice C

## Código de la implementación de la ANCT 1.0

Este apéndice contiene partes del código utilizado para implementar la ANCT 1.0. El primer apartado tiene el código que define la estructura de datos de las tareas topológicas. El segundo apartado contiene el código de implementación de las zonas de un agente.

### La estructura de datos de las tareas topológicas.

A continuación se muestra el código Object Pascal de los tipos de datos: TopologicalTask, TgeneralGoal, TgoalDimension and TconcreteGoal.

```
TgoalDimension = Record // this structure store the information of a general goal
                        relative in a dimension.

                        kindOfSize :integer; // information respect one dimension of
                                                the size relative to the agent.

                        s:relativeTopologicalRelation; // relative topological relation
                                                        between the agent and the object.
end;

TgeneralGoal = Record // this structure implement a general goal

                        pattern:integer; // kind of object
                        zone: integer; // determine the zone of the agent.
                        dimensions: array [1..2] of TgoalDimension;
end;
```

```
TopologicalTask : array [1..z] of TgeneralGoal; // This structure stores general
                                                    //goals to conform a task
TconcreteGoal = Record

    object: RObject;
    zone: integer; // determine the zone of the agent.
    dimensions: array [1..2] of goalDimension;
end;
```

## Código de implementación de la zonas de un agente

La estructura de datos para almacenar la información de las zonas de los agentes es un *array* con una posición para cada zona. En el micromundo, los agentes tienen dos zonas: una es el cuerpo completo del agente y la otra es la boca del agente. Así, la primera posición del *array* almacena información de la zona del cuerpo y la segunda posición almacena sobre la zona de la boca. El código de Object Pascal que se muestra a continuación se usó para implementar las zonas de un agente en el micromundo.

```
zones:array[1..2]of Tcharacteristics; // The data structure to store the information of zones
```

La estructura de datos es creada asignando la información que determina el siguiente código:

```
// egocentric data of the body of the agent
zones[1][0,1].begin :=22;
zones[1][0,1].end :=28;

zones[1][0,2].begin:=0;
zones[1][0,2].end:=0;

// egocentric data of the mouth
zones[2][0,1].begin :=25;//
zones[2][0,1].end :=25;

zones[2][0,2].begin:=0;
zones[2][0,2].end:=0;
```

A los campos de las zonas se les asigna información egocéntrica. Esto permite que la definición de la zona sea válida independientemente de los movimientos que realice el agente.

## Código de implementación de las tareas

El código Object Pascal implementado para la tarea topológica de exploración para los agentes del micromundo es el siguiente:

```
TaskExploring: TopologicalTask;

// fix the pattern which must comply the object
TaskExploring.generalGoal[1].pattern := -1;
//fix the zone of agent which must establish the relation with the object
TaskExploring.generalGoal[1].zone := 1;
//fix the relation of size between the agent and the object
TaskExploring.generalGoal[1].size := 1;
//fix the relation which must me achieved in dimension 1
TaskExploring.generalGoal[1].dimension[1] := Inside;
//fix the relation which must me achieved in dimension 2
TaskExploring.generalGoal[1].dimension[2] := Inside;

// fix the pattern which must comply the object
TaskExploring.generalGoal[2].pattern := -1;
//fix the zone of agent which must establish the relation with the object
TaskExploring.generalGoal[1].zone := 1;
//fix the relation of size between the agent and the object
TaskExploring.generalGoal[2].size := 2;
//fix the relation which must me achieved in dimension 1
TaskExploring.generalGoal[2].dimension[1] := Meet-0;
//fix the relation which must me achieved in dimension 2
TaskExploring.generalGoal[2].dimension[2] := Inside;

// fix the pattern which must comply the object
TaskExploring.generalGoal[3].pattern := -1;
//fix the zone of agent which must establish the relation with the object
TaskExploring.generalGoal[1].zone := 1;
//fix the relation of size between the agent and the object
TaskExploring.generalGoal[3].size := 5;
//fix the relation which must me achieved in dimension 1
TaskExploring.generalGoal[3].dimension[1] := Equal;
//fix the relation which must me achieved in dimension 2
TaskExploring.generalGoal[3].dimension[2] := Inside;
```

El significado del código anterior es el siguiente. El que la meta topológica general en Dimension[2] es lograr la relación “inside” con un objeto de tipo puerta implica que estará siempre garantizado que el agente tendrá un comportamiento de avanzar

como reacción. En el caso de que un obstáculo aparezca en frente del agente, el agente reacciona alineándose con la puerta mediante desplazamientos laterales, lo que a su vez implica que estará evitando los obstáculos en el entorno.

El código Object Pascal para la tarea topológica *situarse para comer* es el siguiente:

```
TaskPositioning: TopologicalTask;

// fix the pattern which must comply the object
TaskPositioning.generalGoal[1].pattern := 4
//fix the zone of agent which must establish the relation with the object
TaskPositioning.generalGoal[1].zone := 2;
//fix the relation of size between the agent and the object
TaskPositioning.generalGoal[1].size := 1
//fix the relation which must be achieved in dimension 1
TaskPositioning.generalGoal[1].dimension[1] := inside
//fix the relation which must be achieved in dimension 2
TaskPositioning.generalGoal[1].dimension[2] := meet-0
```

Si no hay obstáculo entre el agente y el alimento, se combinará la información de la meta general y el objetivo específico para generar una meta específica mediante el siguiente código.

```
specificGoal:TspecificGoal;

specificGoal.object:= object(4).copy; // give the all the information of the object
specificGoal.dimension[1] := inside;
specificGoal.dimension[2] := meet-0;
```

La estructura de datos *concreteGoal.object* contiene toda la información cuantitativa sobre el objeto con el cual la relación espacial se busca establecer. Esta información será usada por la HSCT para calcular la relación topológica relativa entre el agente y el objetivo.

# Apéndice D

## Topología

En este apéndice se incluyen las definiciones y teoremas de topología que han sido usados en la investigación. Se omiten las demostraciones dado que pueden ser consultadas en un texto general de topología [262].

**Definición 18.** *Dado un conjunto,  $X$ , se dice que  $\mathcal{T}$  es una topología definida sobre  $X$  si  $\mathcal{T}$  es una colección de subconjuntos de  $X$  tales que:*

1.  $\emptyset, X \in \mathcal{T}$
2.  $\forall i \in I, U_i \in \mathcal{T} \Rightarrow (\bigcup_i U_i) \in \mathcal{T}$
3.  $U_1 \in \mathcal{T}, U_2 \in \mathcal{T} \Rightarrow (U_1 \cap U_2) \in \mathcal{T}$

A cada elemento de  $\mathcal{T}$  se le llama abierto.

**Definición 19.** *Se denomina<sup>1</sup> espacio topológico al par  $(X, \mathcal{T})$*

**Definición 20.** *Dado un conjunto  $X$ , se dice que  $\mathcal{B}$  es una base si es una colección de subconjuntos que satisface:*

1.  $\forall x \in X, \exists B \in \mathcal{B}$
2.  $\forall x \in B_1 \cap B_2$  con  $B_1, B_2 \in \mathcal{B}, \exists B_3 \in \mathcal{B} : x \in B_3 \subset B_1 \cap B_2$

**Definición 21.** *Dada una base  $\mathcal{B}$  sobre  $X$ , la topología generada por  $\mathcal{B}$ , es aquella tal que,  $\mathcal{U}$  es abierto sí y sólo sí para todo  $x \in \mathcal{U}$  existe  $B \in \mathcal{B}$  tal que  $x \in B \subset \mathcal{U}$ .*

**Proposición 1.** *La topología generada por una base es realmente una topología.*

De alguna manera, una base contiene los ladrillos necesarios para construir cualquier abierto.

**Lema 4.** *Dada una base,  $\mathcal{B}$ , cada uno de sus elementos es abierto en la topología que genera y todo abierto en esa topología se puede escribir como la unión (quizá infinita) de elementos de  $\mathcal{B}$ .*

---

<sup>1</sup>Muchas veces se abrevia  $(X, \mathcal{T})$  por  $X$  si se sobreentiende  $\mathcal{T}$

**D.0.0.0.1 Ejemplo**

La colección de todos los intervalos abiertos  $\mathcal{B}_1 = \{(a, b) : a, b \in \mathbb{R}\}$  es una topología en  $\mathbb{R}$  y se la llama *topología usual*.

Puesto que los intervalos abiertos  $(a, b) = \{x \in \mathbb{R} : a < x < b\}$  generan una topología (la usual). Una idea para generar nuevas topologías es cambiar el significado de “<”

**Definición 22.** Se dice que “ $\leq$ ” es una relación de orden lineal si en un conjunto  $X$  si para todo  $a, b, c \in X$  se cumple

1.  $a \leq a$
2.  $a \leq b, b \leq a \Rightarrow a = b$
3.  $a \leq b, b \leq c \Rightarrow a \leq c$

y además dados  $a, b$  siempre se cumple una de las relaciones  $a < b, b < a$  o  $a = b$ .

Aquí se ha usado  $a < b$  como abreviatura de  $a < b, a \neq b$ . Los intervalos se denotan como en  $\mathbb{R}$ . Los intervalos se denotan como en  $\mathbb{R}$ .

$$(a, b) = \{x : a < x < b\} \quad [a, b] = \{x : a \leq x \leq b\}$$

**D.0.0.0.2 Ejemplo**

Si en el conjunto  $X = \{\clubsuit, \diamond, \heartsuit, \spadesuit\}$  se define la relación de orden  $\clubsuit < \diamond < \heartsuit < \spadesuit$ , entonces

$$(\clubsuit, \heartsuit) = \{\diamond\} \quad (\diamond, \spadesuit) = \{\heartsuit\} \quad (\diamond, \clubsuit) = \emptyset$$

**Definición 23.** Sea  $X$  un conjunto con un orden lineal. Se llama topología del orden a la generada por los intervalos abiertos  $(a, b)$  con  $a, b \in X$  y por  $[m, b)$  y/o  $(a, M]$  si  $X$  tuviera un elemento mínimo,  $m$ , y/o máximo,  $M$ .

**D.0.0.0.3 Ejemplo**

En el ejemplo anterior  $m = \clubsuit$  y  $M = \spadesuit$  y la topología del orden coincide con la discreta, ya que todo elemento de  $X$  es abierto, de hecho es un elemento de la base:

$$\{\clubsuit\} = [\clubsuit, \diamond) \quad \{\diamond\} = (\clubsuit, \heartsuit) \quad \{\heartsuit\} = (\diamond, \spadesuit) \quad \{\spadesuit\} = (\heartsuit, \spadesuit]$$

Dada una topología  $X$  y otra sobre  $Y$  hay una forma canónica de crear una topología sobre el producto cartesiano  $X \times Y$ .

**Definición 24.** Sean  $\mathcal{B}_X$  y  $\mathcal{B}_Y$  bases de topologías sobre  $X$  e  $Y$ , respectivamente. Se llama *topología producto* a la topología sobre  $X \times Y$  generada por

$$\mathcal{B}_X \times \mathcal{B}_Y = \{(B_X, B_Y) : B_X \in \mathcal{B}_X, B_Y \in \mathcal{B}_Y\}$$

Igual que se define la topología producto en  $X \times Y$  se puede definir en  $X_1 \times X_2 \times \dots \times X_n$

En topología se conoce el siguiente lema.

**Lema 5.** Si  $\mathcal{B}_X$  y  $\mathcal{B}'_X$  son bases de una misma topología en  $X$  y  $\mathcal{B}_Y$  y  $\mathcal{B}'_Y$  son bases de una misma topología en  $Y$ , entonces  $\mathcal{B}_X \times \mathcal{B}_Y$  generan la misma topología que  $\mathcal{B}'_X \times \mathcal{B}'_Y$  sobre  $X \times Y$ .



# Apéndice E

## Conferencias y publicaciones realizadas

Las publicaciones que se han realizado hasta el momento de la investigación que se presenta en esta tesis son las siguientes:

### Actas de congresos:

- Miguel-Tomé, S. (2011): General Theory of Exobehaviours: A New Proposal to Unify Behaviors. IWINAC (1) pp. 20-29
- Miguel-Tomé, S. (2013): On the Identification and Establishment of Topological Spatial Relations. IWINAC (2) pp. 461-470
- Miguel-Tomé, S. (2015): Trajectories-State: A New Neural Mechanism to Interpret Cerebral Dynamics. IWINAC (1) pp. 88-97

### Publicaciones en revistas internacionales

- Miguel-Tomé, S. and Fernández-Caballero, A. (2014): On the identification and establishment of topological spatial relations by autonomous systems. Connection Science. 26(3) pp. 261-292.
- Miguel-Tomé, S. (2015): Extensions of the heuristic topological qualitative semantic: Enclosure and fuzzy relations. Robotics and Autonomous Systems 63 pp. 214-218.
- Miguel-Tomé, S. (2017): Navigation through unknown and dynamic open spaces using topological notions. Connection Science. in press.

- Miguel-Tomé,S.(2017): The influence of computational traits on the natural selection of the nervous system. Natural Computing. in press.

# Bibliografía

- [1] Physics of computation. *International Journal of Theoretical Physics* 21(12), 903–903 (1982)
- [2] The differential lambda-calculus. *Theoretical Computer Science* 309(1), 1–41 (2003)
- [3] Special issue: David Marr. *Perception* 41(9), 1009–1162 (2012)
- [4] Â., M.K., et al.: Brainstem circuits underlying the prey-catching behavior of the frog. *Brain Behav Evol* (83), 104–111 (2014)
- [5] A. A. S. Weir, J. Chappell, A.K.: Shaping of hooks in new caledonian crows. *Science* (297), 981 (2002)
- [6] Acar, M., Mettetal, J., van Oudenaarden, A.: Stochastic switching as a survival strategy in fluctuating environments. *Nature Genetics* 40(4), 471–475 (2008)
- [7] Ackley, D.H., et al.: A learning algorithm for boltzmann machines. *Cognitive Science* 9(1), 147–169 (1985)
- [8] Adams, F.: Why we still need a mark of the cognitive. *Cognitive Systems Research* 11(4), 324–331 (2010), special Issue on Extended Mind
- [9] Adams, F., Garrison, R.: The mark of the cognitive. *Minds and Machines* 23(3), 339–352 (2013)
- [10] Addessi, E., et al.: Preference transitivity and symbolic representation in capuchin monkeys (*cebus apella*). *PLOS ONE* 3(6), 1–8 (2008)
- [11] Agerwala, T.: Communication with automata. Hopkins computer research report, John Hopkins University (1974)
- [12] Aho, A., Ullman, J.: Properties of syntax directed translations. *Journal of Computer and System Sciences* 3(3), 319–334 (1969)

- 
- [13] should be unified: comment on Griffiths et al., C.S., et al., M.: Feldman, jerome a. *Trends in cognitive sciences* 14(8), 341 (2010)
- [14] Allen, A., et al.: Feeding-related traits are affected by dosage of the foraging gene in *drosophila melanogaster*. *Genetics* 205(2), 761–773 (2017)
- [15] Allen, J.F.: Maintaining knowledge about temporal intervals. *Communications of the ACM* 26(11), 832–843 (1983)
- [16] Alstott, J., et al.: Modeling the impact of lesions in the human brain. *PLoS Comput. Biol.* 5(6), e1000408 (06 2009)
- [17] Álvarez-Sánchez, J.R., et al.: Reactive navigation in real environments using partial center of area method. *Robotics and Autonomous Systems* 58(12), 1231–1237 (2010)
- [18] Amari, S.I.: Neural theory of association and concept-formation. *Biological Cybernetics* 26(3), 175–185 (1977)
- [19] Ames-III, A.: CNS energy metabolism as related to function. *Brain Research Reviews* 34(1-2), 42–68 (2000)
- [20] Amit, D.J.: *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge University Press (1992)
- [21] Ananthanarayanan, R., Esser, S.K., Simon, H.D., Modha, D.S.: The cat is out of the bag: cortical simulations with  $10^9$  neurons,  $10^{13}$  synapses. In: *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*. pp. 1–12. ACM, New York, NY, USA (2009)
- [22] Anderson, J.R.: *The architecture of cognition*. Harvard University, Cambridge (1983)
- [23] Appel, H.M., Cocroft, R.B.: Plants respond to leaf vibrations caused by insect herbivore chewing. *Oecologia* 175(4), 1257–1266 (2014)
- [24] Apt, K.R.: *From Logic Programming to Prolog*. Prentice-Hall, Inc. (1996)
- [25] Armitage, J., et al.: “Neural Networks” in Bacteria: making connections. *Journal of Bacteriology* 187(1), 26–36 (2005)
- [26] Arrighi, P., Nesme, V., Werner, R.: *One-Dimensional Quantum Cellular Automata over Finite, Unbounded Configurations*, pp. 64–75. Springer Berlin Heidelberg (2008)

- 
- [27] Asher, N., Vieu, L.: Toward a geometry of common sense: a semantics and a complete axiomatization of mereotopology. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence. pp. 846–852 (1995)
- [28] Avery, O., Colin, M., Macleod, M.: Studies on the chemical nature of the substance inducing transformation of pneumococcal types. induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type. The Journal of Experimental Medicine 79(2), 137–158 (1944)
- [29] Avlund, M., Dodd, I.B., Semsey, S., Sneppen, K., Krishna, S.: Why do phage play dice? Journal of Virology 83(22), 11416–11420 (2009)
- [30] Baader, F., Nipkow, T.: Term rewriting and all that. Cambridge University Press (1999)
- [31] Balaban, N., et al.: Bacterial persistence as a phenotypic switch. Science 305(5690), 1622–1625 (2004)
- [32] Baldwin, J.: A new factor in evolution. American Naturalist (30), 441–451 (1896)
- [33] Baluska, F., Mancuso, S.: Root apex transition zone as oscillatory zone. Frontiers in Plant Science 4(354) (2013)
- [34] Baluska, F., et al.: Plant synapses: actin-based domains for cell-to-cell communication. Trends in Plant Science 10(3) (2005)
- [35] Baluska, F., et al.: Vesicular secretion of auxin. Plant Signaling & Behavior 3(4), 254–256 (2008)
- [36] Baluska, F., et al.: The “root-brain” hypothesis of charles and francis darwin. Plant Signaling & Behavior 4(12), 1121–1127 (2009)
- [37] Barraquand, J., Latombe, J.C.: Nonholonomic multibody mobile robots: controllability and motion planning in the presence of obstacles. In: Proceedings of the 1991 IEEE International Conference on Robotics and Automation. vol. 3, pp. 2328–2335 (9-11 1991)
- [38] Bathellier, B., et al.: Discrete neocortical dynamics predict behavioral categorization of sounds. Neuron 76(2), 435–449 (2012)
- [39] Beadle, G., Tatum, E.: Genetic control of biochemical reactions in neurospora. Proceedings of the National Academy of Sciences 27(11), 499–506 (1941)
- [40] Beaumont, H., Gallie, J., Kost, C., Ferguson, G., Rainey, P.: Experimental evolution of bet hedging. Nature (462), 90–93 (2009)

- 
- [41] Beggs, E., Costa, J., Tucker, J.: Three forms of physical measurement and their computability. *Review of Symbolic Logic* 17(1) (2014)
- [42] BEGGS, E.J., COSTA, J.F., TUCKER, J.V.: The impact of models of a physical oracle on computational power. *Mathematical Structures in Computer Science* 22, 853–879 (10 2012)
- [43] Beggs, E.J., Costa, J.F., Tucker, J.V.: Axiomatizing physical experiments as oracles to algorithms. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 370(1971), 3359–3384 (2012)
- [44] Beggs, E., Tucker, J.: Experimental computation of real numbers by newtonian machines. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 463(2082), 1541–1561 (2007)
- [45] Bejan, A.: *Advanced Engineering Thermodynamics*, 3rd Edition. Wiley (2006)
- [46] Bejan, A., Lorente, S.: The constructal law and the evolution of design in nature. *Physics of Life Reviews* 8(3), 209 – 240 (2011)
- [47] Bekenstein, J.D.: Information in the holographic universe. *Scientific American* (2003)
- [48] Bekenstein, J.: Black holes and entropy. *Phys. Rev. D* 7, 2333–2346 (1973)
- [49] Bekenstein, J.: Generalized second law of thermodynamics in black-hole physics. *Phys. Rev. D* 9, 3292–3300 (1974)
- [50] Bekenstein, J.: Holographic bound from second law of thermodynamics. *Physics Letters B* 481(2), 339–345 (2000)
- [51] Bellingham, M.C., Lim, R., Walmsley, B.: Developmental changes in epsc quantal size and quantal content at a central glutamatergic synapse in rat. *The Journal of Physiology* 511(3), 861–869 (1998)
- [52] Ben-Shahar, Y., Robichon, A., Sokolowski, M.B., Robinson, G.E.: Influence of gene action across different time scales on behavior. *Science* 296(5568), 741–744 (2002)
- [53] Beran, M., et al.: Language-trained chimpanzees (pan troglodytes) name what they have seen but look first at what they have not seen. *Psychological Science* 24(5), 660–666 (2013)
- [54] Beran, M., et al.: Comparative cognition: Past, present, and future. *International Journal of Comparative Psychology* 27(1), 1–28 (2014)

- 
- [55] Bertens, L., et al.: Modeling biological gradient formation: combining partial differential equations and petri nets. *Natural Computing* 15(4), 665–675 (2016)
- [56] Bindschaedler, C., et al.: Growing up with bilateral hippocampal atrophy: From childhood to teenage. *Cortex* 47(8), 931–944 (2011)
- [57] Bisio, A., et al.: Free quantum field theory from quantum cellular automata. *Foundations of Physics* 45(10), 1137–1152 (2015)
- [58] Björke, J.: Topological relations between fuzzy regions: derivation of verbal terms. *Fuzzy Sets and Systems* 141(3), 449–467 (2004)
- [59] Bogert, C.M.: Thermoregulation in reptiles, a factor in evolution. *Evolution* 3(3), 195–211 (1949)
- [60] Boghosian, B.M., Taylor, W.: Quantum lattice-gas model for the many-particle schrödinger equation in  $d$  dimensions. *Phys. Rev. E* 57, 54–66 (1998)
- [61] Boisseau, R.P., Vogel, D., Dussutour, A.: Habituation in non-neural organisms: evidence from slime moulds. *Proceedings of the Royal Society of London B: Biological Sciences* 283(1829) (2016)
- [62] Bond, C.: Locomotion and contraction in an asconoid calcareous sponge. *Invertebrate Biology* 132(4), 283–290 (2013)
- [63] Bonnet, J., othes: Amplifying genetic logic gates. *Science* 340(6132), 599–603 (2013)
- [64] Borenstein, J., Koren, Y.: The vector field histogram-fast obstacle avoidance for mobile robots. *IEEE Transactions on Robotics and Automation* 7(3), 278–288 (jun 1991)
- [65] Boring, E.: Intelligence as the tests test it. *New Republic* 36, 35–37 (1923)
- [66] Bouché, N., Fromm, H.: Gaba in plants: just a metabolite? *Trends in Plant Science* 9(3), 110–115 (2004)
- [67] Bournez, O., Campagnolo, M.L.: *A Survey on Continuous Time Computations*, pp. 383–423. Springer New York (2008)
- [68] Bournez, O., Dershowitz, N., Néron, P.: *Axiomatizing Analog Algorithms*, pp. 215–224 (2016)
- [69] Bournez, O., Graña, D., Pouly, A.: Computing with polynomial ordinary differential equations. *Journal of Complexity* 36, 106 – 140 (2016)

- 
- [70] Brachman, R., Levesque, H.: Knowledge Representation and Reasoning. Morgan Kaufmann (2004)
- [71] Brachman, R., Levesque, H.: Knowledge Representation and Reasoning. Elsevier (2004)
- [72] Bradford, M.J., Roff, D.A.: Bet hedging and the diapause strategies of the cricket *allonemobius fasciatus*. *Ecology* 74(4), pp. 1129–1135 (1993)
- [73] Branco, T., Staras, K.: The probability of neurotransmitter release: variability and feedback control at single synapses. *Nature Reviews Neuroscience* 10, 373–383 (2009)
- [74] Bray, D.: Molecular networks: The top-down view. *Science* 301(5641), 1864–1865 (2003)
- [75] Brenner, E.: Plant neurobiology: an integrated view of plant signaling. *Trends in plant science* 11(8), 413–419 (2006)
- [76] Bresadola, M.: Medicine and science in the life of luigi galvani (1737-1798). *Brain Research Bulletin* 46(5), 367–380 (1998)
- [77] Brillouin, L.: Maxwell’s demon cannot operate: Information and entropy. i. *J. Appl. Phys* 22(3), 334–337 (1951)
- [78] Bringsjord, S., Xiao, H.: A refutation of penrose’s gödelian case against artificial intelligence. *Journal of Experimental & Theoretical Artificial Intelligence* 12(3), 307–329 (2000)
- [79] Brodal, P.: *The Central Nervous System*. Oxford University Press (2010)
- [80] Brooks, R.: A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation* 2(1), 14–23 (mar 1986)
- [81] Brooks, R.A.: A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation* 2(1), 14–23 (April 1986)
- [82] Brooks, R.A.: The relationship between matter and life. *Nature* (409-411), 409–411 (January 2001)
- [83] Brooks, R.A.: Elephants don’t play chess. *Robot. Auton. Syst.* 6(1-2), 3–15 (1990)
- [84] Brooks, R.A.: Elephants don’t play chess. *Robotics and Autonomous Systems* 6(1-2), 3–15 (1990)

- 
- [85] Brooks, R.A.: Intelligence without reason. In: Proceedings of the 12th international joint conference on Artificial intelligence. pp. 569–595. Morgan Kaufmann Publishers Inc. (1991)
- [86] Brooks, R.A.: Intelligence without representation. *Artificial Intelligence* 47(1-3) (1991)
- [87] Brooks, R.A.: Intelligence without representation. *Artificial Intelligence* 47(1-3), 139–159 (1991)
- [88] Bullmore, E., Sporns, O.: The economy of brain network organization. *Nature Reviews Neuroscience* 12, 336–349 (2012)
- [89] Butterfield, N.J.: The neoproterozoic. *Current Biology* 25(19), R859–R863 (2015)
- [90] Butz, M., et al.: A model for cortical rewiring following deafferentation and focal stroke. *Frontiers in Computational Neuroscience* 3(10) (2009)
- [91] Ramón y Cajal, S.: The croonian lecture. la fine structure des centres nerveux. *Proc. R. Soc. Lond. B* 55, 444–468 (1894)
- [92] Cannon, W.: *The Wisdom of the Body*. W. W. Norton. (1932)
- [93] Cardelli, L., Zavattaro, G.: On the computational power of biochemistry. In: Horimoto, K., Regensburger, G., Rosenkranz, M., Yoshida, H. (eds.) *Algebraic Biology*, vol. 5147, pp. 65–80 (2008)
- [94] Cartwright, P., et al.: Exceptionally preserved jellyfishes from the middle cambrian. *PLoS ONE* 2(10), e1121 (2007)
- [95] Chase, R.: Structure and function in the cerebral ganglion. *Microscopy Research and Technique* 49(6), 511–520 (2000)
- [96] Chen, Y., et al.: Programmable chemical controllers made from dna. *Nature Nanotechnology* (8), 755–762 (2013)
- [97] Cherniak, C.: Component placement optimization in the brain. *The Journal of Neuroscience* 14(4), 2418–2427 (1994)
- [98] Chestnutt, J., Lau, M., Cheung, G., Kuffner, J.J., Hodgins, J., Kanade, T.: Footstep planning for the Honda ASIMO humanoid. In: Proceedings of the 2005 IEEE International Conference on Robotics and Automation. pp. 629–634 (2005)
- [99] Chiribella, G., et al.: Informational derivation of quantum theory. *Phys. Rev. A* 84, 012311 (2011)

- 
- [100] Choset, H., Lynch, K.M., Hutchinson, S., Kantor, G., Burgard, W., Kavraki, L.E., Thrun, S.: Principles of Robot Motion: Theory, Algorithms, and Implementations. The MIT Press (1995)
- [101] Chou, A., et al.: Search for space-time correlations from the planck scale with the fermilab holometer pp. 1–5 (2015)
- [102] Churchland, P.M., Churchland, P.S.: Could a machine think? *Scientific American* (262), 26–31 (January 1990)
- [103] Clarac, F.: IBRO History of Neuroscience, chap. The History of Reflexes Part 2: From Sherrington to 2004 (2005)
- [104] Clementini, E., Di Felice, P., Hernández, D.: Qualitative representation of positional information. *Artificial Intelligence* 95(2), 317–356 (1997)
- [105] Cohn, A., Renz, J.: Qualitative spatial reasoning. In: *Handbook of Knowledge Representation*. pp. 581–584 (2008)
- [106] Copeland, B.: Hypercomputation. *Minds and Machines* 12(4), 461–502 (2002)
- [107] Costa, J., Loff, B., Mycka, J.: A foundation for real recursive function theory. *Annals of Pure and Applied Logic* 160(3), 255–288 (2009)
- [108] da Costa, N.C.A., Doria, F.A.: Undecidability and incompleteness in classical mechanics. *International Journal of Theoretical Physics* 30(8), 1041–1073 (1991)
- [109] Costello, T., McCarthy, J.: Book review extending mechanics to minds: The mechanical foundations of psychology and economics. *Artificial Intelligence* 170(18), 1237–1238 (2006)
- [110] Craik, K.: *The Nature of Explanation*. Cambridge University Press (1943)
- [111] Crick, F.: The recent excitement about neural networks. *Nature* (337), 129–132 (1989)
- [112] Cronbach, L.: The two disciplines of scientific psychology. *American Psychologist* (12), 671–684 (1957)
- [113] D’Ariano, G.M., Perinotti, P.: Quantum cellular automata and free quantum field theory. *Frontiers of Physics* 12(1), 120301 (2016)
- [114] Darwin, C.: *On the Origin of Species by Means of Natural Selection*. John Murray (1859)

- 
- [115] Darwin, C.: *The descent of man: and selection in relation to sex*. D. Appleton, 2nd ed. edn. (1920), original work published 1871
- [116] Davidson, D.: Truth and meaning. *Synthese* 17(3), 304–323 (1967)
- [117] Davis, M.: The myth of hypercomputation. In: *Alan Turing: Life and Legacy of a Great Thinker*, pp. 195–211. Springer Berlin Heidelberg (2004)
- [118] Davis, M.: Why there is no such discipline as hypercomputation. *Applied Mathematics and Computation* 178(1), 4–7 (2006)
- [119] De Witt, T.J., et al.: Costs and limits of phenotypic plasticity. *Trends in Ecology & Evolution* 13(2), 77–81 (1998)
- [120] Dedrick, D., Trick, L. (eds.): *Computation, Cognition and Pylyshyn*. Massachusetts Institute of Technology (2009)
- [121] Delafontaine, M., et al.: Qualitative relations between moving objects in a network changing its topological relations. *Information Sciences* 178(8), 1997–2006 (2008)
- [122] Delcomyn, F.: *Foundations of neurobiology*. WH Freeman (1999)
- [123] Descartes, R.: *Discurso del método*. Ed. Alfaguara (1981), publicado originalmente en 1637
- [124] Deutsch, D.: Quantum theory, the church-turing principle and the universal quantum computer. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 400(1818), 97–117 (1985)
- [125] Deutsch, D.: *A Computable Universe: Understanding and Exploring Nature as Computation*, chap. What is Computation? (How) Does Nature Compute?, pp. 551–566. World Scientific Publishing Company (2012)
- [126] Dirafzoon, A., Lobaton, E.J.: Topological mapping of unknown environments using an unlocalized robotic swarm. In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. pp. 5545–5551 (2013)
- [127] Dore, J.: Holophrases, speech acts and language universals. *Journal of Child Language* (2), 21–40 (1975)
- [128] Doyle, J.: *Extending Mechanics to Minds*. Cambridge University Press, New York (2006)
- [129] Drayson, Z.: Extended cognition and the metaphysics of mind. *Cognitive Systems Research* 11(4), 367–377 (2010)

- [130] Durr, C., Santha, M.: A decision procedure for well-formed linear quantum cellular automata. *Random Structures and Algorithms* 11(4), 381–394 (1997)
- [131] Duchin, L.: The evolution of articulate speech: comparative anatomy of the oral cavity in pan and homo. *Journal of Human Evolution* 19(6), 687–697 (1990)
- [132] Duckworth, R.A.: The role of behavior in evolution: a search for mechanism. *Evolutionary Ecology* 23(4), 513–531 (2009)
- [133] van Duijn, M.: Phylogenetic origins of biological cognition: convergent patterns in the early evolution of learning. *Interface Focus* 7(3) (2017)
- [134] Durr, C., Santha, M.: A decision procedure for unitary linear quantum cellular automata. *IEEE 37th Annual Symposium on Foundations of Computer Science* pp. 38–45 (1996)
- [135] Dylla, F., Ferrein, A., Lakemeyer, G., Murray, J., Obst, O., Röfer, T., Schiffer, S., Stolzenburg, F., Visser, U., Wagner, T.: Approaching a formal soccer theory from behaviour specifications in robotic soccer. In: *Computer Science and Sports* (2008)
- [136] Dylla, F., Wallgrün, J.: Qualitative spatial reasoning with conceptual neighborhoods for agent control. *Journal of Intelligent and Robotic Systems* 48(1), 55–78 (2007)
- [137] Earman, J., Norton, J.D.: Forever is a day: Supertasks in pitowsky and malament-hogarth spacetimes. *Philosophy of Science* 60(1), 22–42 (1993)
- [138] Eberbach, E., Goldin, D., Wegner, P.: Turing’s ideas and models of computation. In: *Alan Turing: Life and Legacy of a Great Thinker*, pp. 159–194. Springer Berlin Heidelberg (2004)
- [139] van Eck, D., de Jong, H.: Mechanistic explanation, cognitive systems demarcation, and extended cognition. *Studies in History and Philosophy of Science Part A* 59, 11–21 (2016)
- [140] Edis, T., Boudry, M.: Beyond physics? on the prospects of finding a meaningful oracle. *Foundations of Science* 19(4), 403–422 (2014)
- [141] Egenhofer, M., et al.: Topological relations between regions with holes. *Int. Jour. of Geographical Information Systems* (8), 128–142 (1994)
- [142] Eilenberg, S., Lane, M.: General theory of natural equivalences. *Trans. Amer. Math. Soc.* 58, 231–294 (1945)

- 
- [143] Ernst, G., Newell, A.: GPS: a case study in generality and problem solving. Academic Press (1969)
- [144] Etesi, G., Németi, I.: Non-turing computations via malament–hogarth spacetimes. *International Journal of Theoretical Physics* 41(2), 341–370 (2002)
- [145] F., C., et al.: Stable polycomb-dependent transgenerational inheritance of chromatin states in drosophila. *Natura Genetics* (49), 876–886 (2017)
- [146] Fasola, J., Matarić, M.J.: Interpreting instruction sequences in spatial language discourse with pragmatics towards natural human-robot interaction. In: 2014 IEEE International Conference on Robotics and Automation (ICRA). pp. 2720–2727 (2014)
- [147] Fechner, B., Osterloh, A.: A meta-level true random number generator. *Int. J. Crit. Comput.-Based Syst.* 1(1/2/3), 267–279 (2010)
- [148] Fernández-Caballero, A., et al.: Optical flow or image subtraction in human detection from infrared camera on mobile robot. *Robotics and Autonomous Systems* 58(12), 1273–1281 (dec 2010)
- [149] Fernando, C., et al.: Molecular circuits for associative learning in single-celled organisms. *Journal of The Royal Society Interface* 6(34), 463–469 (2009)
- [150] Ferrein, A., Fritz, C., Lakemeyer, G.: Using Golog for deliberation and team coordination in robotic soccer. *Künstliche Intelligenz* 19, 24–30 (2005)
- [151] Feynman, R.: Simulating physics with computers. *International Journal of Theoretical Physics* 21(6-7), 467–488 (1982)
- [152] Feynman, R.: The character of physical law. MIT Press (1964)
- [153] Fikes, R.E., Nilsson, N.J.: Strips: a new approach to the application of theorem proving to problem solving. *Artificial Intelligence* 2(3), 189–208 (1971)
- [154] Fodor, J.A.: Searle on what only brains can do. *Behavioral and Brain Sciences* 3 (1980)
- [155] Fogliaroni, P., Wallgrün, J.O., Clementini, E., Tarquini, F., Wolter, D.: A qualitative approach to localization and navigation based on visibility information. In: Proceedings of the 9th International Conference on Spatial Information Theory. pp. 312–329 (2009)
- [156] Forbus, K.D., Nielsen, P., Faltings, B.: Qualitative spatial reasoning: the clock project. *Artificial Intelligence* 51(1-3), 417–471 (1991)

- [157] Frank, A.: Qualitative spatial reasoning with cardinal directions. In: Proceedings of the 7th Austrian Conference on Artificial Intelligence. pp. 157–167 (1991)
- [158] Fredkin, E.: Digital mechanics. *Physica D* (45), 254–270 (1990)
- [159] Fredkin, E.: A new cosmogony: On the origin of the universe. In: *Physics and Computation, 1992. PhysComp '92., Workshop on.* pp. 116–121. IEEE Press (1992)
- [160] Freksa, C.: Temporal reasoning based on semi-intervals. *Artificial Intelligence* 54(1-2), 199–227 (1992)
- [161] Freksa, C.: Using orientation information for qualitative spatial reasoning. In: *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, vol. 639, pp. 162–178. Springer Berlin Heidelberg (1992)
- [162] Fromm, J., Lautner, S.: Electrical signals and their physiological significance in plants. *Plant, Cell & Environment* 30(3), 249–257 (2007)
- [163] Fugazza, C., et al.: Recall of others' actions after incidental encoding reveals episodic-like memory in dogs. *Current Biology* 26(23), 3209–3213 (2016)
- [164] Fukuda, M., Yamamoto, T., Llinás, R.: The isochronic band hypothesis and climbing fibre regulation of motricity: an experimental study. *European Journal of Neuroscience* 13(2), 315–326 (2001)
- [165] Fussy, S., et al.: Nonlocal computation in quantum cellular automata. *Phys. Rev. A* 48, 3470–3477 (1993)
- [166] G., L., Paul, T.: Computability in Context: Computation and Logic in the Real World, chap. *The Mathematics of Computing between Logic and Physics*, pp. 243–274 (2009)
- [167] Gardner, H.: Models of intelligence: International perspectives, chap. *Three distinct meanings of intelligence*, pp. 133–150. American Psychological Association (2003)
- [168] Galileo, G.: *Il saggiaiore* (1623)
- [169] Galton, F.: *Hereditary genius: an inquiry into its laws and consequences*. Macmillan & Co (1869)
- [170] Gamow, G.: Possible relation between deoxyribonucleic acid and protein structures. *Nature* 173, 318 (1954)
- [171] Gamut, L.T.F.: *Logic, Language, and Meaning, Volume 2: Intensional Logic and Logical Grammar*. University Of Chicago Press (1990)

- 
- [172] Gardner, M.: The fantastic combinations of john conway's new solitaire game "life". *Scientific American* (223), 120–123 (1970)
- [173] Gardner, R.A., Gardner, B.: Teaching sign language to a chimpanzee. *Science* 165(3894), 664–672 (1969)
- [174] Garlick, D.: Understanding the nature of the general factor of intelligence: The role of individual differences in neural plasticity as an explanatory mechanism. *Psychological Review* 109(1), 116–136 (2002)
- [175] Garm, A., Ekström, P., Boudes, M., Nilsson, D.E.: Rhopalia are integrated parts of the central nervous system in box jellyfish. *Cell and Tissue Research* 325(2), 333–343 (2006)
- [176] Gerevini, A., Nebel, B.: Qualitative spatio-temporal reasoning with rcc-8 and allen's interval calculus: Computational complexity. In: *Proceedings of the 15th European Conference on Artificial Intelligence*. pp. 312–316 (2002)
- [177] Gerstner, W., Sprekeler, H., Deco, G.: Theory and simulation in neuroscience. *Science* 338(6103), 60–65 (2012)
- [178] Ghalambor, C., et al.: *Evolutionary Behavioral Ecology*, chap. Behaviour as phenotypic plasticity, pp. 90–107. Oxford University Press (2010)
- [179] Gherardi, M., MandrÅ , S., Bassetti, B., Cosentino Lagomarsino, M.: Evidence for soft bounds in ubuntu package sizes and mammalian body masses. *Proceedings of the National Academy of Sciences* 110(52), 21054–21058 (2013)
- [180] Ghirardi, G., Rimini, A., Weber, T.: A general argument against superluminal transmission through the quantum mechanical measurement process. *Lettere al nuovo cimento* 27(10), 293–298 (1980)
- [181] Ghirardi, G., et al.: Experiments of the epr type involving cp -violation do not allow faster-than-light communication between distant observers. *EPL (Europhysics Letters)* 6(2), 95 (1988)
- [182] Ghosh, A., Pal, N., Pal, S.: Modeling of component failure in neural networks for robustness evaluation: an application to object extraction. *Neural Networks, IEEE Transactions on* 6(3), 648–656 (1995)
- [183] Gierer, A., Meinhardt, H.: A theory of biological pattern formation. *Kybernetik* 12(1), 30–39 (1972)
- [184] Gintis, H.: *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton University Press, revised edition edn. (2014)

- 
- [185] Girard, M.: Interactive design of 3d computer-animated legged animal motion. *IEEE Computer Graphics and Applications* 7(6), 39–51 (june 1987)
- [186] Glover, J., Fritsch, B.: *Encyclopedia of neuroscience*, chap. Brains of primitive chordates, pp. 439–448. Springer-Verlag (2009)
- [187] Golblat, R.: *Topoi: The categorical analysis of logic*. Dover Publications (2006)
- [188] Gottfredson, L. S. Saklofske, D.H.: Intelligence: Foundations and issues in assessment. *Canadian Psychology* 50(3), 183–195 (2009)
- [189] Goyal, R., Egenhofer, M.: The direction-relation matrix: A representation of direction relations for extended spatial objects. In: *UCGIS Annual Assembly and Summer Retreat*. Bar Harbor (1997)
- [190] Grasso, F.W. and Basil, J.: The evolution of flexible behavioral repertoires in cephalopod molluscs. *Brain, Behavior and Evolution* (74), 231–245 (2009)
- [191] Gray, C.M., et al.: Synchronization of oscillatory neuronal responses in cat striate cortex: Temporal properties. *Visual Neuroscience* 8, 337–347 (1992)
- [192] Green, R.M., et al.: Circadian rhythms confer a higher level of fitness to arabidopsis plants. *Plant Physiology* 129(2), 576–584 (2002)
- [193] Green, R., Tingay, S., Wang, Z., Tobin, E.: Circadian rhythms confer a higher level of fitness to arabidopsis plants. *Plant Physiology* 129(2), 576–584 (2002)
- [194] Greenspan, R.: E pluribus unum, ex uno plura: Quantitative and single-gene perspectives on the study of behavior. *Annual Review of Neuroscience* 27(1), 79–105 (2004)
- [195] Grössing, G., Zeilinger, A.: A conservation law in quantum cellular automata. *Physica D: Nonlinear Phenomena* 31(1), 70–77 (1988)
- [196] Grössing, G., Zeilinger, A.: Quantum cellular automata. *Complex systems* 2(2), 197–208 (1988)
- [197] Grössing, G., Zeilinger, A.: Structures in quantum cellular automata. *Physica B+C* 151(1-2), 366–369 (1988)
- [198] Grunfest, H.: Evolution of nervous control from primitive organisms to man, chap. Evolution of conduction in the nervous system, pp. 43–86 (1959)
- [199] Guler, U., et al.: Design of efficient cmos ring oscillator-based random number generator. *International Journal of Electronics* (in press), in press (2017)

- 
- [200] Hagar, A.: *Discrete or Continuous?: The Quest for Fundamental Length in Modern Physics*. Cambridge University Press (2014)
- [201] Hahmann, T., Winter, M., Gruninger, M.: Stonian p-ortholattices: a new approach to the mereotopology RT0. *Artificial Intelligence* 173(15), 1424–1440 (2009)
- [202] Haier, R.J.: Neuro-intelligence, neuro-metrics and the next phase of brain imaging studies. *Intelligence* 37(2), 121–123 (2009)
- [203] Haller, A.: *Commentarii Societatis Regiae Scientiarum Gottingensis*, vol. 2, chap. De partibus corporis humani sensibilibus et irritabilibus, pp. 114–158 (1753)
- [204] Hameroff, S., Penrose, R.: Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. *Mathematics and Computers in Simulation* 40(3), 453–480 (1996)
- [205] Hanken, J., Wake, D.B.: Miniaturization of body size: Organismal consequences and evolutionary significance. *Annual Review of Ecology and Systematics* 24, 501–519 (1993)
- [206] Hari, R., Puce, A.: *MEG-EEG Primer*. Oxford University Press (2017)
- [207] Harlow, D., Hayden, P.: Quantum computation vs. firewalls. *Journal of High Energy Physics* 2013(6), 85 (2013)
- [208] Harmer, S.L.: The circadian system in higher plants. *Annual Review of Plant Biology* 60(1), 357–377 (2009)
- [209] Harnad, S.: The symbol grounding problem. *Physica D: Nonlinear Phenomena* 42(1), 335–346 (1990)
- [210] Hartley, R.: Transmission of information. *Bell System Technical Journal* 7(3), 535–563 (1928)
- [211] Hawking, S.W.: Breakdown of predictability in gravitational collapse. *Phys. Rev. D* 14, 2460–2473 (1976)
- [212] Hawking, S.: Particle creation by black holes. *Communications in Mathematical Physics* 43(3), 199–220 (1975)
- [213] Hayes, K.J., Hayes, C.: The intellectual development of a home-raised chimpanzee. *Proceedings of the American Philosophical Society* 95(2), 105–109 (1951)

- [214] Hazarika, D., Hazarika, D.: Topological relations between fuzzy regions with holes in a general fuzzy topological space. *Journal of Informatics and Mathematical Sciences* 4(2) (2012)
- [215] Hebb, D.: *The organization of behavior: a neuropsychological theory*. Wiley (1949)
- [216] Hedrich, R., Salvador-Recatalá, V., Dreyer, I.: Electrical wiring and long-distance plant communication. *Trends in Plant Science* 21(5), 376–387 (2016)
- [217] Heim, I., Kratzer, A.: *Semantics in Generative Grammar*. Wiley-Blackwell (1998)
- [218] Heinrich, B., Bugnyar, T.: Just how smart are ravens? *Scientific American* 296(4), 64–71 (2007)
- [219] Hempel, C., Oppenheim, P.: Studies in the logic of explanation. *Philosophy of Science* (15), 135–175 (1948)
- [220] Hennessey, T.M., Rucker, W.B., McDiarmid, C.G.: Classical conditioning in paramecia. *Animal Learning and Behavior* 7(4), 417–423 (1979)
- [221] Herrero, J., Mira, J.: Schema: A knowledge edition interface for obtaining program code from structured descriptions of psms: Two case studies. *Applied Intelligence* 10, 139–153 (1999)
- [222] Herrero, J., Mira, J.: Causality levels in schema: a knowledge edition interface. *Software, IEE Proceedings -* 147(5), 193–200 (oct 2000)
- [223] Hilbert, D.: Axiomatisches denken. *Mathematische Annalen* 78, 405–415 (1918)
- [224] Hirsch, M.W.: Convergent activation dynamics in continuous time networks. *Neural Networks* 2(5), 331–349 (1989)
- [225] Hirth, F., et al.: Homeotic gene action in embryonic brain development of drosophila. *Development* (125), 1579–1589 (1998)
- [226] Hjelmfelt, A., Weinberger, E.D., Ross, J.: Chemical implementation of neural networks and turing machines. *Proceedings of the National Academy of Sciences* 88(24), 10983–10987 (1991)
- [227] Hofbauer, J., K., S.: *Evolutionary Games and Population Dynamics*. Cambridge University Press (1998)
- [228] Hofman, M.A.: Energy metabolism, brain size and longevity in mammals. *The Quarterly Review of Biology* 58(4), pp. 495–512 (1983)

- 
- [229] Hogarth, M.L.: Does general relativity allow an observer to view an eternity in a finite time? *Foundations of Physics Letters* 5(2), 173–181 (1992)
- [230] Holland, J.: *Adaptation in Natural and Artificial Systems*. MIT press, first mit press edition edn. (1992)
- [231] Holland, N.D.: Early central nervous system evolution: an era of skin brains? *Nature Review Neuroscie* 4(8), 617–627 (2003)
- [232] Holopainen, J., Blande, J.: *Molecular Plant Volatile Communication*, pp. 17–31. Springer US (2012)
- [233] Hooft, G.: Equivalence relations between deterministic and quantum mechanical systems. *Journal of Statistical Physics* 53(1-2), 323–344 (1988)
- [234] Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* 79(8), 2554–2558 (1982)
- [235] Hornsby, J.: Physicalist thinking and conceptions of behaviour. In: *Subject, Thought, and Context*, pp. 95–116. Oxford University Press (1986)
- [236] Huey, R., et al.: Behavioral drive versus behavioral inertia in evolution: A null model approach. *The American Naturalist* 161(3), 357–366 (2003)
- [237] Hutter, M.: *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Springer, Berlin (2006)
- [238] Huxley, J.: *Evolution: The modern synthesis*. MIT Press, the definitive edition edn. (2010)
- [239] I., T., Y.C., L., S., T.: Predictive behavior within microbial genetic networks. *Science* 320, 1313–1317 (2008)
- [240] Ilachinski, A.: *Cellular Automata: A Discrete Universe*. World Scientific Publishing Co., Inc. (2001)
- [241] Ingram, K., et al.: Context-dependent expression of the foraging gene in field colonies of ants: the interacting roles of age, environment and task. *Proceedings of the Royal Society of London B: Biological Sciences* 283(1837) (2016)
- [242] Ionescu, M., Păun, G., Yokomori, T.: Spiking neural p systems. *Fundamenta Informaticae* 71(2,3), 279–308 (2006)
- [243] Jackson, A.S.: *Analog computation*. McGraw-Hill (1960)

- 
- [244] Jannes, G.: Some comments on “the mathematical universe”. *Foundations of Physics* 39(4), 397–406 (2009)
- [245] Jaynes, E.T.: Information theory and statistical mechanics. *Physical Review* 106, 620–630 (1957)
- [246] Jaynes, E.T.: Information theory and statistical mechanics. ii. *Physical Review* 108, 171–190 (1957)
- [247] Johannsen, W.: The genotype conception of heredity. *The American Naturalist* 45(531), 129–159 (1911)
- [248] Johnson-Laird, P.: Mental models in cognitive science. *Cognitive Science* 4(1), 71–115 (1980)
- [249] Joliot, M., Ribary, U., Llinás, R.: Human oscillatory brain activity near 40 hz coexists with cognitive temporal binding. *Proceedings of the National Academy of Sciences* 91(24), 11748–11751 (1994)
- [250] Jun, Z., Xiaolin, Q.: Topological analysis between bodies with holes. *Int. Jour. of Computer Science and Network Security* 8(5), 167–174 (2008)
- [251] Jung, R., Haier, R.: The parieto-frontal integration theory (p-fit) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences* (30), 135–187 (2007)
- [252] Kala, R., Shukla, A., Tiwari, R.: Dynamic environment robot path planning using hierarchical evolutionary algorithms. *Cybernetics and Systems* 41(6), 435–454 (2010)
- [253] Kala, R., Shukla, A., Tiwari, R.: Robotic path planning using hybrid genetic algorithm particle swarm optimisation. *Int. J. Inf. Commun. Techol.* 4(2/3/4), 89–105 (Aug 2012)
- [254] Kalampokis, A., et al.: Robustness in biological neural networks. *Physica A: Statistical Mechanics and its Applications* 317(3-4), 581–590 (2003)
- [255] Kamon, I., Rimon, E., Rivlin, E.: Tangentbug: a range-sensor based navigation algorithm. *International Journal of Robotics Research* 2, 934–953 (1998)
- [256] de Kamps, M.: Towards truly human-level intelligence in artificial applications. *Cognitive Systems Research* 14(1), 1–9 (2012)
- [257] Karsten, L., Smith, J.: Lattice fermions: Species doubling, chiral invariance and the triangle anomaly. *Nuclear Physics B* 183(1-2), 103–140 (1981)

- 
- [258] Katzenberger, R.J., et al.: A drosophila model of closed head traumatic brain injury. *Proceedings of the National Academy of Sciences* 110(44), E4152–E4159 (2013)
- [259] Kazantsev, V.B., et al.: Self-referential phase reset based on inferior olive oscillator dynamics. *Proceedings of the National Academy of Sciences* 101(52), 18183–18188 (2004)
- [260] Keijzer, F., van Duijn, M., Lyon, P.: What nervous systems do: early evolution, input-output, and the skin brain thesis. *Adaptive Behavior* 21(2), 67–85 (2013)
- [261] Keijzer, F.A.: Evolutionary convergence and biologically embodied cognition. *Interface Focus* 7(3) (2017)
- [262] Kelley, J.: *General Topology*. Springer (1961)
- [263] Kemmerer, D.: The semantics of space: Integrating linguistic typology and cognitive neuroscience. *Neuropsychologia* 44(9), 1607–1621 (2006)
- [264] Kenet, T., et al.: Spontaneously emerging cortical representations of visual attributes. *Nature* (425), 954–956 (2003)
- [265] Khan, A., Schneider, M.: Topological reasoning between complex regions in databases with frequent updates. In: *Proceedings of the 18th SIGSPATIAL*. pp. 380–389. ACM (2010)
- [266] Kim, C., Seong, K.A., Hyung, L.K., Kim, J.O.: Design and implementation of a fuzzy elevator group control system. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 28(3), 277–287 (may 1998)
- [267] Kinnersley, A.M., Turano, F.J.: Gamma aminobutyric acid (gaba) and plant responses to stress. *Critical Reviews in Plant Sciences* 19(6), 479–509 (2000)
- [268] Knoll, A.: The multiple origins of complex multicellularity. *Annual Review of Earth and Planetary Sciences* 39, 217–239 (2011)
- [269] Ko, D.W., Yi, C., Suh, I.H.: Semantic mapping and navigation: A bayesian approach. In: *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. pp. 2630–2636 (2013)
- [270] Köhler, W.: *The Mentality of Apes*. K. Paul, Trench, Trubner and co., ltd, London and New Cork (1925)
- [271] Kohonen, T.: *Associative Memory-A System Theoretical Approach*. Springer-Verlag (1978)

- [272] Kolen, J.F.: Fool's gold: Extracting finite state machines from recurrent network dynamics. In: IN ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 6. pp. 501–508. Morgan Kaufmann (1994)
- [273] Kollar, T., Tellex, S., Roy, D., Roy, N.: Toward understanding natural language directions. In: 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI). pp. 259–266 (March 2010)
- [274] Konrad, Z.: Rechender raum. Elektronische Datenverarbeitung (8), 336–344 (1967)
- [275] Konrad, Z.: Rechender Raum. Friedrich Vieweg & Sohn (1969)
- [276] Kotseruba, I., et al.: A review of 40 years of cognitive architecture research: Focus on perception, attention, learning and applications. CoRR abs/1610.08602 (2016), <http://arxiv.org/abs/1610.08602>
- [277] Kreisel, G.: A notion of mechanistic theory. Synthese 29(1/4), pp. 11–26 (1974)
- [278] Kuba, M., et al.: When do octopuses play? effects of repeated testing, object type, age, and food deprivation on object play in octopus vulgaris. Journal of Comparative Psychology 120(3), 184–190 (2006)
- [279] Kuffner, J.J., Nishiwaki, K., Kagami, S., Inaba, M., Inoue, H.: Footstep planning among obstacles for biped robots. In: Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. vol. 1, pp. 500–505 (2001)
- [280] Kuipers, B.: The spatial semantic hierarchy. Artificial Intelligence 119(1-2), 191–233 (2000)
- [281] Kussell, E., Leibler, S.: Phenotypic diversity, population growth, and information in fluctuating environments. Science 309(5473), 2075–2078 (2005)
- [282] Kwalski, R., Sergot, M.: A logic-based calculus of events. New Generation Computing 4(1), 67–95 (1986)
- [283] Laghi, T.: Sulla insensibilitÃ ed irritabilitÃ Halleriana, vol. 2, chap. De sensitivitate atque irritabilitate Halleriana, pp. 326–345. Corciolani ed Eredi Colli (1757)
- [284] Laland, K., et al.: An introduction to niche construction theory. Evolutionary Ecology 30(2), 191–202 (2016)
- [285] Lane, S.M.: Categories for the working mathematician. Springer (1998)

- 
- [286] Langley, P., Laird, J.E., Rogers, S.: Cognitive architectures: Research issues and challenges. *Cognitive Systems Research* 10(2), 141–160 (June 2009)
- [287] Lapedra, O., et al.: Behavioural changes and the adaptive diversification of pigeons and doves. *Proceedings of the Royal Society of London B: Biological Sciences* 280(1755) (2013)
- [288] Larkum, M.E., Zhu, J.J., Sakmann, B.: A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature* 398, 338–341 (1999)
- [289] Larson, G.E., Saccuzzo, D.P.: Cognitive correlates of general intelligence: Toward a process theory of g. *Intelligence* 13(1), 5–31 (1989)
- [290] Laughlin, S.B., Sejnowski, T.J.: Communication in neural networks. *Science* 301(5641), 1870–1874 (2003)
- [291] Lavalle, S.M.: *Planning Algorithms*. Cambridge University Press (2006)
- [292] Lea, R.N., Hoblit, J., Jani, Y.: Fuzzy logic based robotic arm control. In: *Second IEEE International Conference on Fuzzy Systems*. vol. 1, pp. 128–133 (1993)
- [293] Lee, T.D.: Difference equations as the basis of fundamental physical theories. In: *Old and New Problems in Fundamental Physics* (1984)
- [294] Lee, T.: Can time be a discrete dynamical variable? *Physics Letters B* 122(3-4), 217–220 (1983)
- [295] Lee, T.: Discrete mechanics. In: *How Far Are We from the Gauge Forces*, pp. 15–114. Springer US (1985)
- [296] Lefebvre, L.: Chapter 19 - primate encephalization. In: Hofman, M.A., Falk, D. (eds.) *Evolution of the Primate Brain, Progress in Brain Research*, vol. 195, pp. 393–412 (2012)
- [297] Lenton, T.M., et al.: Co-evolution of eukaryotes and ocean oxygenation in the neoproterozoic era. *Nature Geoscience* (7), 257–265 (2014)
- [298] Levesque, H.J., Reiter, R., Lespérance, Y., Lin, F., Scherl, R.B.: Golog: a logic programming language for dynamic domains. *The Journal of Logic Programming* 31(1-3), 59–83 (1997)
- [299] Levitt, T.S., Lawton, D.T.: Qualitative navigation for mobile robots. *Artificial Intelligence* 44(3), 305–360 (1990)
- [300] Levy, W.B., Baxter, R.A.: Energy-efficient neuronal computation via quantal synaptic failures. *The Journal of Neuroscience* 22(11), 4746–4755 (2002)

- [301] Lewis, E.: A gene complex controlling segmentation in drosophila. *Nature* 276(5688), 565–570 (1978)
- [302] Lewontin, R.: The units of selection. *Annual Review of Ecology and Systematics* 1, 1–18 (1970)
- [303] Lewontin, R.: Evolution from molecules to men, chap. Gene, organism, and environment, pp. 273–285. Cambridge University Press (1983)
- [304] Leys, S.P.: Elements of a ‘nervous system’ in sponges. *Journal of Experimental Biology* 218(4), 581–591 (2015)
- [305] Leys, S., Mackie, G.: Electrical recording from a glass sponge. *Nature* 387, 29–30 (1997)
- [306] Li, S., Wang, H.: Rcc8 binary constraint network can be consistently extended. *Artificial Intelligence* 170(1), 1–18 (2006)
- [307] Li, S., Ying, M.: Generalized region connection calculus. *Artificial Intelligence* 160(1), 1–34 (2004)
- [308] Li, Y., Li, S.: A fuzzy sets theoretic approach to approximate spatial reasoning. *IEEE Transactions on Fuzzy Systems* 12(6), 745–754 (2004)
- [309] Lin, H.: Architectural Design of Multi-Agent Systems: Technologies and Techniques. University of Houston-Downtown (2007)
- [310] Liu, W., Zhang, X., Li, S., Ying, M.: Reasoning about cardinal directions between extended objects. *Artificial Intelligence* 174(12-13), 951–983 (2010)
- [311] Liverant, S.: Intelligence: A concept in need of re-examination. *Journal of Consulting Psychology* 24(2), 101–110 (1960)
- [312] Llinas, R.: The intrinsic electrophysiological properties of mammalian neurons: Insights into central nervous system function. *Science* (242), 1654–1664 (1988)
- [313] Llinás, R., et al.: Dendritic spikes and their inhibition in alligator purkinje cells. *Science* 160(3832), 1132–1135 (1968)
- [314] Llinás, R.: Mindwaves, chap. “Mindness” as a Functional State of the Brain, pp. 339–358. Oxford (1987)
- [315] Llinás, R.: I of the vortex: from neurons to self. MIT press (2001)

- 
- [316] Llinás, R.R., et al.: Gamma-band deficiency and abnormal thalamocortical activity in P/Q-type channel mutant mice. *Proceedings of the National Academy of Sciences* 104(45), 17819–17824 (2007)
- [317] Lloyd, S.: Ultimate physical limits to computation. *Nature* 406(6799), 1047–1054 (2000)
- [318] LM, A.: Molecular computation of solutions to combinatorial problems. *Science* 266(5187), 1021–1024 (1994)
- [319] Lorenz, K.: The comparative method in studying innate behavior patterns. In: Press, C.U. (ed.) *Physiological Mechanisms of Animal Behavior: Symposia of the Society of Experimental Biology*. pp. 221–268 (1950)
- [320] Love, P.J., Boghosian, B.M.: From dirac to diffusion: Decoherence in quantum lattice gases. *Quantum Information Processing* 4(4), 335–354 (2005)
- [321] Lozano-Perez, T.: Spatial planning: A configuration space approach. *IEEE Transactions on Computers* C-32(2), 108–120 (feb 1983)
- [322] Lumelsky, J.V., Stepanov, E.A.: Path planning strategies for a point mobile automaton moving amidst unknown obstacles of arbitrary shape. *Algorithmica* 17(9), 403–430 (1987)
- [323] Lyon, P.: Developing scaffolds in evolution, culture and cognition, chap. Stress in mind: a stress response hypothesis cognitive cognition, pp. 171–190. MIT Press (2013)
- [324] Lyon, P.: The cognitive cell: bacterial behavior reconsidered. *Frontiers in Microbiology* 6, 264 (2015)
- [325] M., E., Miller: Intelligence and brain myelination: A hypothesis. *Personality and Individual Differences* 17(6), 803–832 (1994)
- [326] Mackie, G.O.: Neuroid conduction and the evolution of conducting tissues. *The quarterly review of biology* 45(4), 319–332 (1970)
- [327] Mackie, G.: Central neural circuitry in the jellyfish *aglantha*. *Neurosignals* 13(1-2), 5–19 (2004)
- [328] Macphail, E.M.: *Brain and Intelligence in Vertebrates*. Oxford University press (1982)
- [329] Macphail, E.M., Barlow, H.B.: Vertebrate intelligence: The null hypothesis [and discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 308(1135), 37–51 (1985)

- 
- [330] Magaña-Loiza, O., et al.: Exotic looped trajectories of photons in three-slit interference. *Nature Communications* (7), 13987 (2016)
- [331] Magnasco, M.O.: Chemical kinetics is turing universal. *Physical Review Letters* 78(6), 1190–1193 (1997)
- [332] de Mairan, J.: *Observation botanique. Histoire de l'Academie royale des sciences* pp. 35–36 (1729)
- [333] Makarenko, V., Llinás, R.: Experimentally determined chaotic phase synchronization in a neuronal system. *Proceedings of the National Academy of Sciences* 95(26), 15747–15752 (1998)
- [334] Mandadapu, K.K., et al.: Mechanics of torque generation in the bacterial flagellar motor. *Proceedings of the National Academy of Sciences* 112(32), E4381–E4389 (2015)
- [335] Manzano, M.: *Extensions of First Order Logic*. Cambridge University Press, Cambridge (1996)
- [336] Manzano, M.: *Model Theory*. Oxford University Press (1999)
- [337] Marcillac, F., et al.: A single mutation alters production and discrimination of drosophila sex pheromones. *Proceedings of the Royal Society of London B: Biological Sciences* 272(1560), 303–309 (2005)
- [338] Margolus, N.: Physics-like models of computation. *Physica D* 10(1-2), 81–95 (1984)
- [339] Marijuán, P.C., Navarro, J., del Moral, R.: On prokaryotic intelligence: Strategies for sensing the environment. *Biosystems* 99(2), 94–103 (2010)
- [340] Mark F, B., Connors, B.W., Paradiso, M.A.: *Neuroscience: Exploring the Brain*. Lippincott Williams & Wilkins (2007)
- [341] Marr, D.: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt and Co., Inc., New York, NY, USA (1982)
- [342] Martin, P., Bateson, P.: *Measuring Behaviour. An Introductory Guide*. Cambridge, 3rd edn. (2009)
- [343] Masi, E., et al.: Spatiotemporal dynamics of the electrical network activity in the root apex. *Proceedings of the National Academy of Sciences* 106(10), 4048–4053 (2009)

- 
- [344] Masoud, A.: Kinodynamic motion planning. *IEEE Robotics Automation Magazine* 17(1), 85–99 (March 2010)
- [345] Matarić, M.: Behavior-based robotics as a tool for synthesis of artificial behavior and analysis of natural behavior. *Trends in Cognitive Sciences* 2(3), 82–86 (1998)
- [346] Maturana, H., Varela, F.: *The Tree of Knowledge: The Biological Roots of Human Understanding*. Shambhala (1992)
- [347] Maturana, H.R.: The organization of the living: A theory of the living organization. *International Journal of Man-Machine Studies* 7(3), 313–332 (1975)
- [348] Maynard Smith, J.: *Evolution and the Theory of Games*. Cambridge University Press (1984)
- [349] Mayor, J., et al.: Connectionism coming of age: legacy and future challenges. *Frontiers in Psychology* 5, 187 (2014)
- [350] Mayr, E.: *Animal species and evolution*. Belknap Press (1963)
- [351] Mayr, E.: The objects of selection. *Proceedings of the National Academy of Sciences* 94(6), 2091–2094 (1997)
- [352] Mazzoni, P., Andersen, R.A., Jordan, M.I.: A more biologically plausible learning rule for neural networks. *Proceedings of the National Academy of Sciences* 88(10), 4433–4437 (1991)
- [353] McCarthy, J., Hayes, P.J.: Some philosophical problems from the standpoint of artificial intelligence. In: *Readings in Nonmonotonic Reasoning*. pp. 26–45 (1987)
- [354] McCarthy, J., other: A proposal for the dartmouth summer research project on artificial intelligence (August 1955)
- [355] McCarthy, J., Hayes, P.J.: Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence* (4), 463–502 (1969)
- [356] McClelland, M., Campbell, M., Estlin, T.: Qualitative relational mapping and navigation for planetary rovers. *Robotics and Autonomous Systems* 83, 73–86 (2016)
- [357] McClung, C.R.: Plant circadian rhythms. *The Plant Cell* 18(4), 792–803 (2006)
- [358] McClung, C.: Plant circadian rhythms. *Plant Cell* 18(4), 792–803 (2006)
- [359] McCulloch, W., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* (5), 115 – 133 (1943)

- [360] McCulloch, W., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* 5(4), 115–133 (1943)
- [361] McDermott, D., Davis, E.: Planning routes through uncertain territory. *Artificial Intelligence* 22(1), 107–156 (1984)
- [362] McKenney, M., Praing, R., Schneider, M.: Deriving topological relationships between simple regions with holes. In: *Headway in Spatial Data Handling*, pp. 521–531. Springer Berlin Heidelberg (2008)
- [363] Mendel, G.: Versuche über pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn* (4), 3–47 (1866)
- [364] Mery, F., et al.: Natural polymorphism affecting learning and memory in drosophila. *Proceedings of the National Academy of Sciences* 104(32), 13051–13055 (2007)
- [365] Meyer, D.A.: From quantum cellular automata to quantum lattice gases. *Journal of Statistical Physics* 85(5), 551–574 (1996)
- [366] Meyer, D.A.: On the absence of homogeneous scalar unitary cellular automata. *Physics Letters A* 223(5), 337–340 (1996)
- [367] Meyers, E.M., Freedman, D.J., Kreiman, G., Miller, E.K., Poggio, T.: Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology* 100(3), 1407–1419 (2008)
- [368] Miguel-Tomé, S.: Hacia una teoría de unificación para los comportamientos cognitivos (2008)
- [369] Miguel-Tomé, S.: Sobre la unificación de los comportamientos cognitivos. In: *Avances en Informática y Automática, 3<sup>o</sup> Workshop*. Departamento de Informática y Automática, Universidad de Salamanca (2009)
- [370] Miguel-Tomé, S.: General theory of exobehaviours: A new proposal to unify behaviors. In: *Foundations on Natural and Artificial Computation, Lecture Notes in Computer Science*, vol. 6686, pp. 20–29. Springer Berlin / Heidelberg (2011)
- [371] Miguel-Tomé, S.: Principios matemáticos del pensamiento natural: Teoría cognitiva de condiciones de verdad. Gráficas Quintanilla (2006)
- [372] Miguel-Tomé, S.: Estudio de algoritmos de semántica cualitativa aplicados a la navegación. Master's thesis, Universidad de Castilla la Mancha, Ingeniería Informática (September 2008)

- 
- [373] Miguel-Tomé, S.: On the identification and establishment of topological spatial relations. In: IWINAC 2013. vol. 2, pp. 461–470. Springer-Verlag Berlin Heidelberg (2013)
- [374] Miguel-Tomé, S.: Extensions of the heuristic topological qualitative semantic: Enclosure and fuzzy relations. *Robotics and Autonomous Systems* (63), 214–218 (2014)
- [375] Miguel-Tomé, S.: Extensions of the heuristic topological qualitative semantic: Enclosure and fuzzy relations. *Robot. Auton. Sys.* (63), 214–218 (2015)
- [376] Miguel-Tomé, S.: Trajectories-state: A new neural mechanism to interpretate cerebral dynamics. In: *Artificial Computation in Biology and Medicine, Lecture Notes in Computer Science*, vol. 9107, pp. 88–97. Springer International Publishing (2015)
- [377] Miguel-Tomé, S.: The influence of computational traits on the natural selection of the nervous system. *Natural Computing* (in press), in press (2017)
- [378] Miguel-Tomé, S.: Navigation through unknown and dynamic open spaces using topological notions. *Connection Science* (in press), in press (2017)
- [379] Miguel-Tomé, S., Fernández-Caballero, A.: On the identification and establishment of topological spatial relations by autonomous systems. *Connection Science* 26(3), 261–292 (2014)
- [380] Miller, A.: Ecologic factors that accelerate formation of races and species of terrestrial vertebrates. *Evolution* 10(3), 262–277 (1956)
- [381] Miller, G.A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63(2), 81–97 (1956)
- [382] Miller, G.A.: The cognitive revolution: a historical perspective. *Trends in cognitive sciences* 7(3), 141–144 (2003)
- [383] Minsky, M.: Steps toward artificial intelligence. *Proceedings of the IRE* 49(1), 8–30 (1961)
- [384] Minsky, M.: *The Society of Mind*. Simon and Schuster, New York (1986)
- [385] Minsky, M., Papert, S.: *Perceptrons*. MIT press (1969)
- [386] Minsky, M.: A framework for representing knowledge. Memo 306, MIT-AI Laboratory (1974)

- [387] Minsky, M.: Cellular vacuum. *International Journal of Theoretical Physics* 21(6-7), 537–551 (1982)
- [388] Minsky, M.L.: *Computation: Finite and Infinite Machines*. Prentice-Hall, Inc. (1967)
- [389] Mira, J., Delgado, A.: Always trying to write an equation for the brain. In: *Artificial Neural Networks, Lecture Notes in Computer Science*, vol. 540, pp. 93–100. Springer Berlin / Heidelberg (1991)
- [390] Mira, J., Delgado, A.E.: In: *Proc of the Seventh International Congress of Cybernetic and Systems*, chap. Some comments on the antropocentric viewpoint in the neurocybernetic methodology., pp. 891–895 (1987)
- [391] Mira, J., Delgado, A.E.: In: *Autonomous robotic systems*, chap. Where is knowledge in robotics? some methodological issues on symbolic and connectionist perspectives of AI, pp. 3–34. Physica-Verlag (2003)
- [392] Mira, J., Delgado, A.: Neural modeling in cerebral dynamics. *Biosystems* 71(1-2), 133–144 (2003)
- [393] Mira, J., Delgado, A.: Sensory representation spaces in neuroscience and computation. *Neurocomputing* 72(4), 793–805 (2009)
- [394] Mira, J.M., García, A.E.: On how the computational paradigm can help us to model and interpret the neural function. *Natural Computing* 6(3), 211–240 (2007)
- [395] Mira, J.M.: Symbols versus connections: 50 years of artificial intelligence. *Neurocomputing* 71(4-6), 671–680 (2008)
- [396] Mitchell, A., et al.: Adaptive prediction of environmental changes by microorganisms. *Nature* (460), 220–224 (2009)
- [397] Monk, T.: *The Evolutionary Origin of Nervous Systems and Implications for Neural Computation*. Ph.d. thesis, University of Otago (2014)
- [398] Monk, T., Paulin, M.G.: Predation and the origin of neurones. *Brain Behav Evol* 84, 246–261 (2014)
- [399] Monk, T., Paulin, M.G., Green, P.: Ecological constraints on the origin of neurones. *Journal of Mathematical Biology* 71(6), 1299–1324 (2015)
- [400] Moore, C.: Recursion theory on the reals and continuous-time computation. *Theoretical Computer Science* 162(1), 23–44 (1996)

- 
- [401] Moran, N.: The evolutionary maintenance of alternative phenotypes 139, 971–989 (1992)
- [402] Moratz, R.: Representing relative direction as a binary relation of oriented points. In: Proceedings of the 2006 Conference on ECAI 2006. pp. 407–411 (2006)
- [403] Moratz, R., et al.: Qualitative spatial reasoning about line segments. In: ECAI 2000. pp. 234–238. IOS Press (2000)
- [404] Moratz, R., et al.: Qualitative spatial reasoning about relative position. In: Spatial Cognition III, vol. 2685, pp. 385–400. Springer Berlin Heidelberg (2003)
- [405] Moreno, H., et al.: Synaptic transmission block by presynaptic injection of oligomeric amyloid beta. Proceedings of the National Academy of Sciences 106(14), 5901–5906 (2009)
- [406] Moroz, L.L., Kohn, A.B.: Unbiased view of synaptic and neuronal gene complement in ctenophores: Are there pan-neuronal and pan-synaptic genes across metazoa? Integrative and Comparative Biology 55(6), 1028–1049 (2015)
- [407] Moroz, L.: On the independent origins of complex brains and neurons. Brain Behaviour Evolution (74), 177–190 (2009)
- [408] Moroz, L., Kohn, A.: Independent origins of neurons and synapses: insights from ctenophores. Philosophical Transactions of the Royal Society of London B: Biological Sciences 371(1685), 1–14 (2016)
- [409] Mueller, E.: Commonsense Reasoning: An Event Calculus Based Approach. Morgan Kaufmann, 2nd edition edn. (2015)
- [410] Mycka, J., Costa, J.: Real recursive functions and their hierarchy. Journal of Complexity 20(6), 835–857 (2004)
- [411] MYERS, C.S.: I. instinct and intelligence. British Journal of Psychology, 1904-1920 3(3), 209–218 (1910)
- [412] Navlakha, S., Barth, A., Bar-Joseph, Z.: Decreasing-rate pruning optimizes the construction of efficient and robust distributed networks. PLoS Comput Biol 11(7) (07 2015)
- [413] Newell, A.: The knowledge level. AI Magazine 2(2), 1–20 (1981)
- [414] Newell, A.: Unified Theories of Cognition. Harvard University Press, Cambridge, Massachusetts (1990)

- [415] Newell, A., Shaw, J., Simon, H.: Report on a general problem-solving program. In: Proceedings of the I.C.I.P. pp. 256–264 (1959)
- [416] Newell, A., Simon, A.H.: GPS, a program that simulates human thought pp. 279–293 (1963)
- [417] Newell, A.: You can't play 20 questions with nature and win. In: Visual Information Processing, pp. 283–308. Academic Press (1973)
- [418] Newell, A.: Reflections on the knowledge level. *Artificial Intelligence* 59(1-2), 31–38 (1993)
- [419] Newell, A., Simon, H.A.: Computer science as empirical inquiry: symbols and search. *Commun. ACM* 19(3), 113–126 (1976)
- [420] Newman, W., Hogan, N.: High speed robot control and obstacle avoidance using dynamic potential functions. In: Proceedings of the 1987 IEEE International Conference on Robotics and Automation. vol. 4, pp. 14–24 (mar 1987)
- [421] Newton, I.: Sir Isaac Newton's Mathematical Principles of Natural Philosophy and His System of the World. University of California Press, translated into english by Andrew Motte 1729, reed.1947 edn. (1934)
- [422] Newton, I.: Sir Isaac Newton's Mathematical Principles of Natural Philosophy and His System of the World., vol. 2. University of California Press, translated into english by Andrew Motte 1729, reed.1947 edn. (1934)
- [423] Nickel, M.: Evolutionary emergence of synaptic nervous systems: what can we learn from the non-synaptic, nerveless porifera? *Invertebrate Biology* 129(1), 1–16 (2010)
- [424] Nickel, M., et al.: The contractile sponge epithelium sensu lato – body contraction of the demosponge *tethya wilhelma* is mediated by the pinacoderm. *Journal of Experimental Biology* 214(10), 1692–1698 (2011)
- [425] Nielsen, H., Ninomiya, M.: Absence of neutrinos on a lattice: (i). proof by homotopy theory. *Nuclear Physics B* 185(1), 20–40 (1981)
- [426] Niklas, K.J., Newman, S.A.: The origins of multicellular organisms. *Evolution & Development* 15(1), 41–52 (2013)
- [427] Nilsson, D.E., Gislén, L., Coates, M.M., Skogh, C., Garm, A.: Advanced optics in a jellyfish eye. *Nature* 435, 201–205 (2005)

- 
- [428] Niven, J.E., Laughlin, S.B.: Energy limitation as a selective pressure on the evolution of sensory systems. *Journal of Experimental Biology* 211(11), 1792–1804 (2008)
- [429] Nămeti, I., Dăvid, G.: Relativistic computers and the turing barrier. *Applied Mathematics and Computation* 178(1), 118–142 (2006)
- [430] N.S., R., Metzler, J.: Mental rotation of three-dimensional objects. *Science* 171(3972), 701–703 (1971)
- [431] Nüslein-Volhard, C., Wieschaus, E.: Mutations affecting segment number and polarity in drosophila. *Nature* 287(30), 795 (1980)
- [432] Odling-Smee, J.: Evolution, the Extended Synthesis, chap. Niche Inheritance, pp. 175–203. MIT press (2010)
- [433] Okasha, S.: *Evolution and the Levels of Selection*. Oxford University Press (2006)
- [434] Okubo, F., Yokomori, T.: The computational capability of chemical reaction automata. *Natural Computing* 15(2), 215–224 (2016)
- [435] Oliveira, A.G., et al.: Circadian control sheds light on fungal bioluminescence. *Current Biology* 25(7), 964–968 (2015)
- [436] Omlin, C.: A Field Guide to Dynamical Recurrent Networks, chap. Understanding and Explaining DRN Behavior, pp. 207–227. Wiley-IEEE Press (2001)
- [437] Ophir, A., et al.: Variation in neural vlar predicts sexual fidelity and space use among male prairie voles in semi-natural settings. *Proceedings of the National Academy of Sciences* 105(4), 1249–1254 (2008)
- [438] O’Reilly, R.C.: Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation* 8(5), 895–938 (Jul 1996)
- [439] Osborne, K., et al.: Natural behavior polymorphism due to a cgmp-dependent protein kinase of drosophila. *Science* 277(5327), 834–836 (1997)
- [440] Oyarce, P., Gurovich, L.: Evidence for the transmission of information through electric potentials in injured avocado trees. *Journal of Plant Physiology* 168(2), 103–108 (2011)
- [441] Pang, T., Maslov, S.: Universal distribution of component frequencies in biological and technological systems. *Proceedings of the National Academy of Sciences* 110(15), 6235–6239 (2013)

- [442] Pantin, C.: The origin of the nervous system. *Publicazioni della Stazione Zoologica di Napoli* (28), 171–181 (1956)
- [443] Parker, G.: *Primitive nervous systems*. Lippincott (1919)
- [444] Passano, L.M.: Primitive nervous systems. *PNAS* 50(2), 306–313 (1963)
- [445] Paulsen, O., Heggelund, P.: Quantal properties of spontaneous epscs in neurones of the guinea-pig dorsal lateral geniculate nucleus. *The Journal of Physiology* 496(3), 759–772 (1996)
- [446] Păun, G.: *Membrane Computing: An Introduction*. Springer Berlin Heidelberg (2002)
- [447] Pavón, J., et al.: Development of intelligent multi-sensor surveillance systems with agents. *Robotics and Autonomous Systems* 55(12), 892–903 (December 2007)
- [448] Pearce, J.: *Animal Learning and Cognition: An Introduction*. Psychology press, 3rd edition edn. (2008)
- [449] Penrose, R.: *The Emperor's New Mind*. Oxford University Press, Oxford (1989)
- [450] Pepperberg, I.: Cognition in the african grey parrot: Preliminary evidence for auditory/vocal comprehension of the class concept. *Animal Learning & Behavior* 11(2), 179–185 (1983)
- [451] Pepperberg, I.: Cognition in an african gray parrot (*psittacus erithacus*): Further evidence for comprehension of categories and labels. *Journal of comparative psychology* 104(1), 41–52 (1990)
- [452] Pepperberg, I.: Further evidence for addition and numerical competence by a grey parrot (*psittacus erithacus*). *Animal Cognition* 15(4), 711–717 (2012)
- [453] Pepperberg, I.: Animal language studies: What happened? *Psychonomic Bulletin & Review* 24(1), 181–185 (2017)
- [454] Pepperberg, I.: Talking with alex: Logic and speech in parrots. *Scientific American* 9(4), 60–65 (1998)
- [455] Pepperberg, I.M.: Functional vocalizations by an african grey parrot (*psittacus erithacus*). *Zeitschrift für Tierpsychologie* 55(2), 139–160 (1981)
- [456] Pepperberg, I.M., Nakayama, K.: Robust representation of shape in a grey parrot (*psittacus erithacus*). *Cognition* 153, 146–160 (2016)

- 
- [457] Pérez-Delgado, C.A., Cheung, D.: Local unitary quantum cellular automata. *Phys. Rev. A* 76, 032320 (2007)
- [458] Petri, C.A.: Communication with automata. Tech. Rep. RADC-TR-65-377, New York: Griffiss Air Force Base (1966)
- [459] Piaget, J.: *Origins of intelligence in the child*. Routledge and Kegan Paul. (1936)
- [460] Piaget, J.: *La psychologie de l'intelligence*. Routledge and Kegan Paul. (1947)
- [461] Pigliucci, M. and Müller, G. (ed.): *Evolution, the Extended Synthesis*. MIT press (2010)
- [462] Pigliucci, M.: Ecological and evolutionary genetics of arabidopsis. *Trends in Plant Science* 3(12), 485–489 (1998)
- [463] Pinker, S.: The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences* 107(Supplement 2), 8993–8999 (2010)
- [464] Pitowsky, I.: The physical church-turing thesis and physical computational complexity. *Iyyun* 5(39), 81–99 (1990)
- [465] Plotkin, H.C.: The role of behavior in evolution, chap. Behavior and evolution, pp. 1–17. MIT Press (1988)
- [466] Polack, C., McConnell, B., Miller, R.: Associative foundation of causal learning in rats. *Learning and Behavior* 41(1), 25–41 (2013)
- [467] Polilov, A.: Anatomy of the smallest of the coleoptera, feather-winged beetles from tribe nanosellini (coleoptera, ptiliidae) and limits to insect miniaturization. *Entomological Review* 88, 26–33 (2008)
- [468] Polilov, A.A.: The smallest insects evolve anucleate neurons. *Arthropod Structure and Development* 41(1), 29 – 34 (2012)
- [469] Pour-El, M.B., Zhong, N.: The wave equation with computable initial data whose unique solution is nowhere computable. *Mathematical Logic Quarterly* 43(4), 499–509 (1997)
- [470] Pour-El, M.B., Richards, I.: The wave equation with computable initial data such that its unique solution is not computable. *Advances in Mathematics* 39(3), 215–239 (1981)
- [471] Preston, J., Bishop, M. (eds.): *Views into the chinese room*. Oxford University Press (2002)

- 
- [472] Putnam, H.: Representation and Reality. MIT Press (1988)
- [473] Pylyshyn, Z.W.: Computational and Cognition: Toward a Foundation for Cognitive Science. MIT press (1984)
- [474] Qian, L., Soloveichik, D., Winfree, E.: Efficient turing-universal computation with dna polymers. In: DNA Computing and Molecular Programming, vol. 6518, pp. 123–140. Springer International Publishing (2011)
- [475] Rabin, M.O.: Probabilistic automata. Information and Control 6(3), 230 – 245 (1963)
- [476] Ramesh, S.A., et al.: Gaba signalling modulates plant growth by directly regulating the activity of plant-specific anion transporters. Nature Communications 6(7879) (2015)
- [477] Randell, D.A., Cui, Z., Cohn, A.G.: A spatial logic based on regions and connection. In: Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning. pp. 165–176 (1992)
- [478] Rashevsky, N.: Mathematical Biophysics: Physico-Mathematical Foundations of Biology. The University of Chicago press (1938)
- [479] Ratcliff, W., et al.: Experimental evolution of multicellularity. Proceedings of the National Academy of Sciences 109(5), 1595–1600 (2012)
- [480] Remolina, E., Kuipers, B.: Towards a general theory of topological maps. Artificial Intelligence 152(1), 47–104 (2004)
- [481] Renard, E., et al.: Origin of the neuro-sensory system: new and expected insights from sponges. Integrative Zoology 4(3), 294–308 (2009)
- [482] Renn, S., Schumer, M.: Genetic accommodation and behavioural evolution: insights from genomic studies. Animal Behaviour 85(5), 1012–1022 (2013)
- [483] Renz, J.: Maximal tractable fragments of the region connection calculus: a complete analysis. In: Proceedings of the 16th International Joint Conference on Artificial intelligence. pp. 448–454 (1999)
- [484] Renz, J.: Qualitative Spatial Reasoning with Topological Information. Springer (2002)
- [485] Renz, J., Nebel, B.: On the complexity of qualitative spatial reasoning: a maximal tractable fragment of the region connection calculus. Artificial Intelligence 108(1-2), 69–123 (1999)

- 
- [486] Rescorla, R.A.: Pavlovian conditioning and its proper control procedures. *Psychological Review* (74), 71–80 (1967)
- [487] Richter, J., et al.: Pull or push? octopuses solve a puzzle problem. *PLOS ONE* 11(3), 1–16 (03 2016)
- [488] Richter, S., Werner, R.F.: Ergodicity of quantum cellular automata. *Journal of Statistical Physics* 82(3), 963–998 (1996)
- [489] Roback, A.A.: Intelligence and behavior. *Psychological Review* 29(1), 54–62 (1922)
- [490] Roberts, A.: *Plasmodesmal Structure and Development*, pp. 1–32 (2007)
- [491] Rosen, R.: Church’s thesis and its relation to the concept of realizability in biology and physics. *The bulletin of mathematical biophysics* 24(4), 375–393 (1962)
- [492] Rosenblatt, F.: The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review* 65(6), 386–408 (1958)
- [493] Roskam, J., Brakefield, P.: Seasonal polyphenism in bicyclus (lepidoptera: Satyridae) butterflies: different climates need different cues. *Biological Journal of the Linnean Society* 66(3), 345–356 (1999)
- [494] Ross, T.: *Fuzzy logic with engineering applications*. John Wiley and Sons, 3rd edn. (2010)
- [495] Rumelhart, D. E. David E. Rumelhart, J.L.M.: *Parallel Distributed Processing. Exploration of the Microstructure of Cognition*, chap. Learning Internal Representation by Error Propagation, pp. 318–362. MIT Press (1986)
- [496] Russell, R., Bab-Hadiashar, A., Shepherd, R.L., Wallace, G.G.: A comparison of reactive robot chemotaxis algorithms. *Robotics and Autonomous Systems* 45(2), 83–97 (2003)
- [497] Sacerdott, E.D.: Planning in a hierarchy of abstraction spaces. In: *Proceedings of the 3rd International Joint Conference on Artificial Intelligence*. pp. 412–422 (1973)
- [498] Salmon, W.: *Four decades of scientific explanation*. University of Pittsburgh Press (1989)
- [499] Samsonovich, A.: Toward a unified catalog of implemented cognitive architectures. In: *Proceeding of the 2010 conference on biologically inspired cognitive architectures*. pp. 195–244 (2010)

- 
- [500] Santhanam, T.S., Tekumalla, A.R.: Quantum mechanics in finite dimensions. *Foundations of Physics* 6(5), 583–587 (1976)
- [501] Santhanam, T.: Quantum mechanics in a finite number of dimensions. *Physica A: Statistical Mechanics and its Applications* 114(1), 445–447 (1982)
- [502] Santos, E.S.: Probabilistic turing machines and computability. *Proceedings of the American Mathematical Society* 22(3), 704–710 (1969)
- [503] Satterlie, R.: Control of swimming in jellyfish: A comparative story. *Canadian Journal of Zoology* (80), 1654–1669 (2002)
- [504] Satterlie, R.A.: Do jellyfish have central nervous systems? *The Journal of Experimental Biology* 214(8), 1215–1223 (2011)
- [505] Schlieder, C.: Representing visible locations for qualitative navigation. In: *Qualitative Reasoning and Decision Technologies*. pp. 523–532 (1993)
- [506] Schlinger, H.: The myth of intelligence. *The physiological record* (53), 15–32 (2003)
- [507] Schmid, U., et al.: The challenge of complexity for cognitive systems. *Cognitive Systems Research* 12, 211–218 (2011)
- [508] Schmidhuber, J.: Gödel machines: self-referential universal problem solvers making provably optimal self-improvements. In: *Tecnicl Report IDSIA-19-03*. pp. 256–264 (2003)
- [509] Schmidt-Nielsen, K.: *Scaling: Why is Animal Size So Important?* Cambridge University Press (1984)
- [510] Schneider, M., Behr, T.: Topological relationships between complex spatial objects. *ACM Trans. Database Syst.* 31(1), 39–81 (2006)
- [511] Schockaert, S., De Cock, M., Kerre, E.E.: Spatial reasoning in a fuzzy region connection calculus. *Artificial Intelligence* 173(2), 258–298 (2009)
- [512] Schreiber, G.: *Knowledge Engineering and Management: The CommonKADS Methodology*. The MIT Press, New York (1999)
- [513] Schrödinger, E.: *What is life?* Cambridge University Press (1944)
- [514] Scialdone, A., et al.: Arabidopsis plants perform arithmetic division to prevent starvation at night. *Elife* (2:e00669) (2013)

- 
- [515] Scialdone, A., Howard, M.: How plants manage food reserves at night: quantitative models and open questions. *Frontiers in Plant Science* 6(204) (2015)
- [516] Scivos, A., Nebel, B.: The finest of its class: The natural point-based ternary calculus  $\mathcal{LR}$  for qualitative spatial reasoning. In: *Spatial Cognition IV. Reasoning, Action, Interaction*, vol. 3343, pp. 283–303. Springer Berlin Heidelberg (2005)
- [517] Searle, J.: Is the brain's mind a computer program? *Scientific American* (262), 26–31 (January 1990)
- [518] Searle, J.R.: Minds, brains, and programs. *Behavioral and Brain Sciences* 3, 417–457 (1980)
- [519] Searle, J.R.: *Mind Design II*, chap. Mind, brain and programs. The MIT press (1997)
- [520] Seeley, T.: *The Wisdom of the Hive: the social physiology of honey bee colonies*. Harvard University Press (1996)
- [521] Sekar, K., et al.: Cortical response tracking the conscious experience of threshold duration visual stimuli indicates visual perception is all or none. *Proceedings of the National Academy of Sciences* 110(14), 5642–5647 (2013)
- [522] Serrano, J., et al.: Cognitive? science? *Foundations of Science* 19(2), 115–131 (2014)
- [523] Seymour, R.S., Schultze-Motel, P.: Thermoregulating lotus flowers. *Nature* (383), 305 (1996)
- [524] Shannon, C.: A mathematical theory of communication. *Bell System Technical Journal*, The 27(3), 379–423 (1948)
- [525] Shannon, C.: A mathematical theory of communication. *Bell System Technical Journal*, The 27(4), 623–656 (1948)
- [526] Shannon, C.E.: Mathematical theory of the differential analyzer. *Journal of Mathematics and Physics* 20(1-4), 337–354 (1941)
- [527] Shapiro, A.: Newton's experimental philosophy. *Early Science and Medicine* 9(3), 185–217 (2004)
- [528] Shi, C., Wang, Y., Yang, J.: Online topological map building and qualitative localization in large-scale environment. *Robotics and Autonomous Systems* 58(5), 488–496 (2010)

- [529] Sibaoka, T.: Rapid plant movements triggered by action potentials. *The botanical magazine = Shokubutsu-gaku-zasshi* 104(1), 73–95 (1991)
- [530] Siegelmann, H.T.: *Neural Networks and Analog Computation: Beyond the Turing Limit* (Progress in Theoretical Computer Science). Birkhäuser Boston (1998)
- [531] Silvertown, J., Gordon, D.: A framework for plant behavior. *Annual Review of Ecology and Systematics* 20, 349–366 (1989)
- [532] Simon, H.A.: Artificial intelligence: an empirical science. *Artificial Intelligence* 77(1), 95–127 (1995)
- [533] Skinner, B.: *The Behavior of Organisms*. Appleton-Century (1938)
- [534] Smith, C., Pivovarova, N., Reese, T.: Coordinated feeding behavior in *trichoplax*, an animal without synapses. *PLoS ONE* 10(9), e0136098 (09 2015)
- [535] Smith, C., et al.: Tactical multimodal signalling in birds: facultative variation in signal modality reveals sensitivity to social costs. *Animal Behaviour* 82(3), 521–527 (2011)
- [536] Smith, C., et al.: Novel cell types, neurosecretory cells, and body plan of the early-diverging metazoan *trichoplax adhaerens*. *Current Biology* 24(14), 1565–1572 (2014)
- [537] Sobek, R.P., Chatila, R.G.: Integrated planning and execution control for an autonomous mobile robot. *Artificial Intelligence in Engineering* 3(2), 103 – 113 (1988)
- [538] Sober, E.: The adaptive advantage of learning and a priori prejudice. *Ethology and Sociobiology* 15(1), 55–56 (1994)
- [539] Sollberger, A.: *Biological Rhythm Research*. Elsevier (1965)
- [540] Soloveichik, D., Seelig, G., Winfree, E.: Dna as a universal substrate for chemical kinetics. *Proceedings of the National Academy of Sciences* 107(12), 5393–5398 (2010)
- [541] Spencer, Q.: Do newton’s rules of reasoning guarantee truth ... must they? *Studies in History and Philosophy of Science Part A* 35(4), 759–782 (2004)
- [542] Sperry, R.W.: Effect of 180 degree rotation of the retinal field on visuomotor coordination. *Journal of Experimental Zoology* 92(3), 263–279 (1943)

- 
- [543] Steels, L.: Symbols and Embodiment: Debates on meaning and cognition, chap. The symbol grounding problem has been solved, so what's next?, pp. 223–244. Oxford University Press (2008)
- [544] Stell, J.G.: Boolean connection algebras: a new approach to the region-connection calculus. *Artificial Intelligence* 122(1-2), 111–136 (2000)
- [545] Stenhouse, D.: A general theory for the evolution of intelligent behaviour. *Nature* 208, 815 (1965)
- [546] Stenhouse, D.: *The Evolution of Intelligence: A General Theory and Some of Its Implications*. Barners & Noble (1973)
- [547] Sternberg, J., Pretz, J. (eds.): *Cognition and Intelligence: Identifying mechanisms of the mind*. Cambridge University Press (2005)
- [548] Sternberg, R.J.: The concept of intelligence and its role in lifelong learning and success. *American Psychologist* 52(10), 1030–1037 (1997)
- [549] Sternberg, R., et al. (eds.): *Models of intelligence: International perspectives*. American Psychological Association (2003)
- [550] Stevenson, H.W.: Piaget, behavior theory, and intelligence. *Monographs of the Society for Research in Child Development* 27(2), 113–126 (1962)
- [551] Stosiek, C., et al.: In vivo two-photon calcium imaging of neuronal networks. *Proceedings of the National Academy of Sciences* 100(12), 7319–7324 (2003)
- [552] Sukhov, V., Nerush, V., Orlova, L., Vodeneev, V.: Simulation of action potential propagation in plants. *Journal of Theoretical Biology* 291, 47–55 (2011)
- [553] Sultan, S.: Phenotypic plasticity for plant development, function and life history. *Trends in Plant Science* 5(12), 537–542 (2000)
- [554] Susskind, L.: The world as a hologram. *Journal of Mathematical Physics* 36(11), 6377–6396 (1995)
- [555] Szilard, L.: On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Systems Research and Behavioral Science* 9(4), 301–310 (1964)
- [556] Szudzik, M.P.: *Some Applications of Recursive Functionals to the Foundations of Mathematics and Physics*. Ph.d. thesis, Carnegie Mellon University (December 2010)

- [557] 't Hooft, G.: Dimensional reduction in quantum gravity (1993)
- [558] 't Hooft, G.: Duality between a deterministic cellular automaton and a bosonic quantum field theory in 1+1 dimensions. *Foundations of Physics* 43(5), 597–614 (2013)
- [559] 't Hooft, G.: *The Cellular Automaton Interpretation of Quantum Mechanics*. Springer International Publishing (2016)
- [560] Tang, X., othes: Topological relations between fuzzy regions in a fuzzy topological space. *International Journal of Applied Earth Observation and Geoinformation* 12, Supplement 2, S151 – S165 (2010)
- [561] Tarsky, A.: *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Oxford: Clarendon Press (1956)
- [562] Taylor, P.: A lambda calculus for real analysis. *Journal of Logic and Analysis* 2(5), 1–115 (2010)
- [563] Tegmark, M.: The importance of quantum decoherence in brain processes. *Physical Review E* 61, 4194–4206 (2000)
- [564] Tegmark, M.: The mathematical universe. *Foundations of Physics* 38(2), 101–150 (2008)
- [565] Tellex, S., et al.: Approaching the symbol grounding problem with probabilistic graphical models. *AI Magazine* 32(4), 64–76 (2011)
- [566] Terrace, H.: *Nim*. Eyre Methuen (1979)
- [567] Thielscher, M.: *Reasoning Robots: The Art and Science of Programming Robotic Agents*. Springer Netherlands (2005)
- [568] Thomas, M., Karmiloff-Smith, A.: Models of intelligence: International perspectives, chap. Connectonist models of development, developmental disorders, and individual differences, pp. 133–150. American Psychological Association (2003)
- [569] Thome, C., et al.: Axon-carrying dendrites convey privileged synaptic input in hippocampal neurons. *Neuron* 83(6), 1418–1430 (2014)
- [570] Toates, F.: The interaction of cognitive and stimulus response processes in the control of behaviour. *Neuroscience & Biobehavioral Reviews* 22(1), 59–83 (1997)
- [571] Toffoli, T.: Cellular automata mechanics. Tech. Rep. Tech. Rep. No. 208, The University of Michigan (November 1977)

- 
- [572] Toffoli, T.: Computation and construction universality of reversible cellular automata. *Journal of Computer and System Sciences* 15(2), 213–231 (1977)
- [573] Toffoli, T.: Cellular automata as an alternative to (rather than an approximation of) differential equations in modeling physics. *Physica D: Nonlinear Phenomena* 10(1-2), 117–127 (1984)
- [574] Tong, D.: Is quantum reality analog after all? *Scientific American* (December 2012)
- [575] Trewavas, A.: Aspects of plant intelligence: an answer to firm. *Annals of Botany* 93(4), 353 (2004)
- [576] Trewavas, A.: What is plant behaviour? *Plant, Cell & Environment* 32(6), 606–616 (2009)
- [577] Trewavas, A.: Aspects of plant intelligence. *Annals of botany* 92(1), 1–20 (2003)
- [578] Tribus, M., McIrvine, E.C.: Energy and information. *Scientific American* 225(3), 180 (1971)
- [579] Trillas, E., et al.: On conjectures in orthocomplemented lattices. *Artificial Intelligence* 117(2), 255–275 (2000)
- [580] Trillas, E., Garc a-Honrado, I., Pradera, A.: Consequences and conjectures in preordered sets. *Information Sciences* 180(19), 3573–3588 (2010)
- [581] Tsotsos, J.K.: Behaviorist intelligence and the scaling problem. *Artificial Intelligence* 75(2), 135 – 160 (1995)
- [582] Tsuda, I.: Toward an interpretation of dynamic neural activity in terms of chaotic dynamical systems. *Behavioral and Brain Sciences* 24(5), 793– 810 (2001)
- [583] Tsuda, I.: Hypotheses on the functional roles of chaotic transitory dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 19(1), 015113 (2009)
- [584] Turing, A.M.: On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society* s2-42(1), 230–265 (1936)
- [585] Turing, A.M.: Systems of logic based on ordinals. *Proceedings of the London Mathematical Society* s2-45(1), 161–228 (1939)
- [586] Turing, A.M.: The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 237(641), 37–72 (1952)

- [587] Turing, A.: Computing machinery and intelligence. *Mind* 59(236), 433–460 (1950)
- [588] Uludag, K., Ugurbil, K., Berliner, L. (eds.): *fMRI: From Nuclear Spins to Brain Functions (Biological Magnetic Resonance)*. Springer (2015)
- [589] van der Maas, H., et al.: Intelligence is what the intelligence test measures. seriously. *Journal of Intelligence* 2(1), 12–15 (2014)
- [590] Varela, F.J.: *Principles of biological autonomy*. New York : North Holland (1979)
- [591] Vauclair, J., Perret, P.: The cognitive revolution in europe: taking the developmental perspective seriously. *Trends in Cognitive Sciences* 7(7), 284–285 (2003)
- [592] Vogeley, K., et al.: Neural correlates of first-person perspective as one constituent of human self-consciousness. *J. Cognitive Neuroscience* 16(5), 817–827 (2004)
- [593] Vogeley, K., Fink, G.R.: Neural correlates of the first-person-perspective. *Trends in Cognitive Sciences* 7(1), 38–42 (2003)
- [594] Volkov, A., Foster, J., Markin, V.: Signal transduction in *mimosa pudica*: biologically closed electrical circuits. *Plant, Cell & Environment* 33(5), 816–827 (2010)
- [595] Volkov, A.G., Markin, V.S.: Active and passive electrical signaling in plants. *Progress in Botany* 76, 143–176 (2015)
- [596] de Waal, F.a.: Towards a bottom-up perspective on animal and human cognition. *Trends in Cognitive Sciences* 14(5), 201–207 (2010)
- [597] de Waal, Frans B.M.; Ferrari, P.F.: Trends in cognitive sciences. Towards a bottom-up perspective on animal and human cognition 14(5), 201–207 (2010)
- [598] Waddington, C.: The epigenotype. *Endeavour* 1, 18–20 (1942)
- [599] Wagner, T., Huebner, K.: An egocentric qualitative spatial knowledge representation based on ordering information for physical robot navigation. *Lecture Notes in Artificial Intelligence* 3276, 134–149 (2005)
- [600] Wagner, T., Huebner, K., Visser, U.: An extended panorama: efficient qualitative spatial knowledge representation for highly dynamic environments. In: *Workshop on Issues in Designing Physical Agents for Dynamic Real-Time Environments*. pp. 109–116 (2003)
- [601] Wall, J., et al.: Human brain plasticity: an emerging view of the multiple substrates and mechanisms that cause cortical changes and related sensory dysfunctions after injuries of sensory inputs from the body. *Brain Research Reviews* 39(2-3), 181–215 (2002)

- 
- [602] Walter, S., Kästner, L.: The where and what of cognition: The untenability of cognitive agnosticism and the limits of the motley crew argument. *Cognitive Systems Research* 13(1), 12–23 (2012)
- [603] Walters, R.: *Categories and Computer Science*. Cambridge University Press (1992)
- [604] Wang, Y., Chiew, V.: On the cognitive process of human problem solving. *Cognitive Systems Research* 11(1), 81–92 (2010)
- [605] Wasmann, E.: *Comparative studies in the psychology of ants and of higher animals*. St. Louis, B. Herder, authorized english version of the 2d german ed. edn. (1905)
- [606] Watrous, J.: On one-dimensional quantum cellular automata. *IEEE 36th Annual Symposium on Foundations of Computer Science* pp. 528–537 (1995)
- [607] Watson, J., Crick, F.: A structure for deoxyribose nucleic acid. *Nature* (171), 737–738 (1953)
- [608] Van de Weghe, N., et al.: Representing moving objects in computer-based expert systems: the overtake event example. *Expert Systems with Applications* 29(4), 977–983 (November 2005)
- [609] Van de Weghe, N., et al.: A qualitative trajectory calculus as a basis for representing moving objects in geographical information systems. *Control and Cybernetics* 35(1), 97–119 (2006)
- [610] Van de Weghe, N., et al.: How to handle incomplete knowledge concerning moving objects. In: *Proceedings of the International Workshop on Behaviour Monitoring and Interpretation*. pp. 91–101 (2007)
- [611] Wehner, R.: Sensory physiology: brainless eyes. *Nature* 435(7039), 157–159 (2005)
- [612] Welch, P.D.: *Discrete Transfinite Computation*. Springer International Publishing (2015)
- [613] Welch, P.: Characteristics of discrete transfinite time turing machine models: Halting times, stabilization times, and normal form theorems. *Theoretical Computer Science* 410(4), 426–442 (2009)
- [614] Welch, P.D.: *Discrete Transfinite Computation Models*, pp. 371–410. Imperial College Press (2012)

- 
- [615] Wenke, D., et al.: Cognition and Intelligence: Identifying the Mechanisms of the Mind, chap. Complex Problem Solving and Intelligence: Empirical Relation and Causal Direction, pp. 160–187. Cambridge University Press (2005)
- [616] Weyl, H.: Quantenmechanik und gruppentheorie. *Zeitschrift für Physik* 46(1), 1–46 (1927)
- [617] Wharton, K.: Questioning the Foundations of Physics, chap. The Universe is not a computer, pp. 177–189. Springer (2015)
- [618] Wheeler, J.A.: Pregeometry: Motivations and Prospects. In: *Quantum Theory and Gravitation*, vol. 1, p. 1 (1980)
- [619] Wheeler, J.A.: Complexity, Entropy, and the Physics of Information, chap. Information, Physics, Quantum: The Search for Links. Addison-Wesley (1990)
- [620] Wiener, N.: *Cybernetics: Or Control and Communication in the Animal and the Machine*. MIT Press (1948)
- [621] Wiener, N., Rosenblueth, A.: The mathematical formulation of the problem of conduction of impulses in a network of connected excitable elements, specifically in cardiac muscle. *Arch. Inst. Cardiol. México* 16, 205–265 (1946)
- [622] Williams, S.E.: Comparative sensory physiology of the droseraceae the evolution of a plant sensory system. *Proceedings of the American Philosophical Society* 120(3), 187–204 (1976)
- [623] Wilm, E.: *The theories of instinct: a study in the history of psychology*. Yale University Press (1925)
- [624] Winskel, G.: *The formal semantics of programming languages*. The MIT press (1993)
- [625] Winter, Y.: *Elements of Formal Semantics*. Edinburgh University Press (2016)
- [626] Wolfram, S.: Statistical mechanics of cellular automata. *Rev. Mod. Phys.* 55, 601–644 (1983)
- [627] Wolfram, S.: Undecidability and intractability in theoretical physics. *Phys. Rev. Lett.* 54(8), 735–738 (1985)
- [628] Wolter, D., Lee, J.H.: Qualitative reasoning with directional relations. *Artificial Intelligence* 174, 1498–1507 (2010)
- [629] Wund, M.: Assessing the impacts of phenotypic plasticity on evolution. *Integrative and Comparative Biology* 52(1), 5 (2012)

- 
- [630] Yan, X., et al.: Research progress on electrical signals in higher plants. *Progress in Natural Science* 19(5), 531–541 (2009)
- [631] Yang, R., Lenaghan, S.C., Zhang, M., Xia, L.: A mathematical model on the closing and opening mechanism for venus flytrap. *Plant Signaling & Behavior* 5(8), 968–978 (2010)
- [632] Yokawa, K., et al.: Binary decisions in maize root behavior: Y-maze system as tool for unconventional computation in plants. *Plant Signaling & Behavior* 10(5-6), 381–390 (2014)
- [633] Yoshida, K.: Challenge: Concept of system life and its application to robotics. *Robotics and Autonomous Systems* 58(7), 833–839 (2010)
- [634] Zadeh, L.: Toward human level machine intelligence - is it achievable? the need for a paradigm shift. *Computational Intelligence Magazine, IEEE* 3(3), 11 –22 (august 2008)
- [635] Zhao, D., et al.: High-resolution non-contact measurement of the electrical activity of plants in situ using optical recording. *Scientific Reports* 5, 13425 (2015)
- [636] Zimmerman, H., Neuneier, R.: A Field Guide to Dynamical Recurrent Networks, chap. *Neural Network Architectures for the Modeling of Dynamic Systems*, pp. 311–350. Wiley-IEEE Press (2001)
- [637] Zylberberg, A., et al.: The human turing machine: a neural framework for mental programs. *Trends in Cognitive Sciences* 15(7), 293–300 (2011)
- [638] Zylberberg, A., Dehaene, S., Roelfsema, P.R., Sigman, M.: The human turing machine: a neural framework for mental programs. *Trends in cognitive sciences* 15(7), 293–300 (2011)