

Universidad Nacional Autónoma de México

Agentes computacionales
cognitivos reactivos
en la modelación de la
conducta visuomotora
de la rana

Tesis de posgrado para obtener el grado de
Maestro en Ciencias de la Computación

presenta:

Virginia Angélica García Vega

México, 1993

TESIS CON
FALLA DE ORIGEN



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Introducción

"Si el punto es descubrir principios muy generales y abstractos de la organización cognoscitiva ... ¿por qué no hacer una criatura totalmente cognitiva, una iguana [rana] marciana de tres ruedas [dos motores], digamos, y un nicho ambiental para que ella interaccione? Pienso que tal proyecto pudiera enseñarnos mucho sobre los principios profundos de la psicología cognitiva humana..."

Daniel Dennett, 1978.



El epígrafe contiene las ideas fundamentales de la tesis y de su objetivo: descubrir *principios cognitivos* que permitan el diseño de arquitecturas computacionales para robots. La figura 1 muestra el mapa de capítulos que expresan las posibles trayectorias de búsqueda que se pueden elegir y la trayectoria seleccionada por nosotros para encontrar dichos principios. El epígrafe implica experimentación con robots construidos *ex-profeso*.

Nótese en la figura 1 que para llegar a cualquier parte experimental de las tres posibles, se debe recorrer una trayectoria que implica una serie de compromisos conceptuales y de arquitectura inevitables.

La tesis, en su parte experimental no intenta la construcción de una *Rana computatrix* [Arbib85], apegada a la conducta global de estos animales ni a la fisiología de los núcleos del sistema nervioso central de la rana [CervantesF-LaraR-ArbibM85], ni mucho menos a la estructura neuronal de ellas --de ahí la denominación de *marciana*.. Pretende la construcción de un robot navegante efectivo, con objetivos asociado y que ejecute, parcialmente por lo menos, una conducta biológica como la de la rana. Tal simplificación nos ha permitido acceder a los principios generales y abstractos a los que se refiere [Dennett78].

Antes de proseguir, una nota sobre el uso indistinto que en esta tesis hacemos de agente y robot:

Dada la naturaleza de la tesis, nos hemos concentrado en entidades libres que siempre tienen una acción potencial sobre el *mundo*, a partir de su atención perceptual hacia él. Esto los califica de robots [Brooks91] . Sin embargo, estamos conscientes que en la literatura del tema se califica de agentes a entidades autónomas que no necesariamente atienden al *mundo*, ni lo modifican.

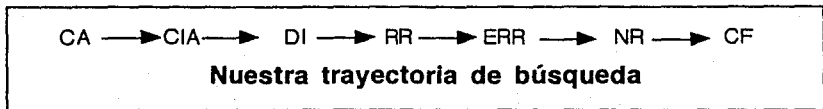
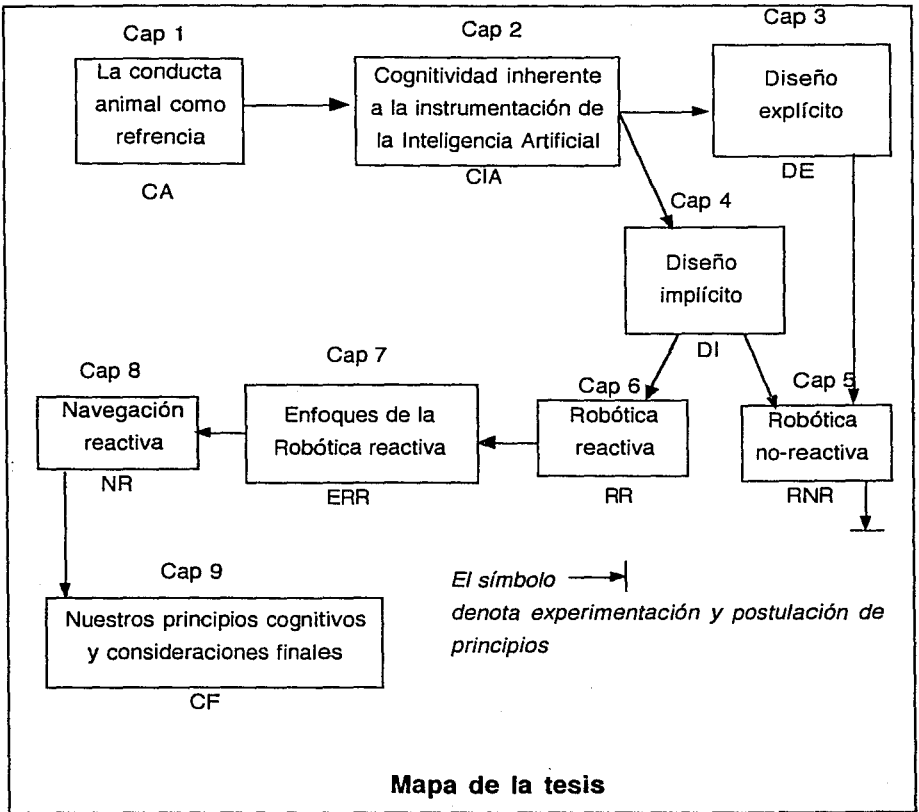


Figura 1

Puesto que la conducta biológica es el paradigma fundamental de esta tesis, en este capítulo se revisarán:

a) los conocimientos etológicos, psicológicos y neuroetológicos más importantes que han influido el diseño de robots inteligentes.

b) Dos de las actitudes inquisitivas fundamentales relativas a la conducta de la rana.



Inteligencia Artificial ha dependido, por lo menos implícita y en ocasiones explícitamente, de la comprensión popular de la autonomía del comportamiento humano y animal.

En este capítulo, se revisan ciertos trabajos en etología, psicología y neurociencia que muestran cuán deficiente puede ser nuestra comprensión diaria ("folk psychology") del comportamiento para los fines de la Inteligencia Artificial y su "implementación" robótica

La conducta animal como un paradigma que permite desarrollar autonomía en robots

¿Cuáles son las *características biológicas* que podrían proporcionar conocimiento en la investigación sobre autonomía?

Se propone que algunas conductas primitivas animales se usen como patrones de comportamiento que se espera puedan surgir en la experimentación con robots. Esto es, las estrategias de control conocidas en los seres vivos proveerán una poderosa inspiración para la construcción de arquitecturas de robots, particularmente para aquéllos sometidos a ambientes no-estructurados y dinámicamente cambiantes como en los que están los seres vivos. Además esto podrá quizá, facilitar la interacción segura entre el hombre y la máquina. Nótese que este acercamiento al *control* de los seres vivos es independiente de nuestra comprensión popular de su conducta.

La investigación sobre *autonomía etológica* ha inspirado poco a la investigación sobre autonomía robótica, puesto que menos esfuerzos de investigación se han hecho sobre ella que sobre el desarrollo de sistemas que manejan clases específicas de problemas. Además estos últimos desarrollos no tienen mucho en común con el control con el que los seres vivos ejecutan las mismas tareas.

1 La conducta animal como referencia

Al manejar clases específicas de problemas, frecuentemente se viola el siguiente principio biológico:

Para que un ser vivo pueda actuar autónomamente sobre un rango amplio de tareas y ambientes, debe ser capaz de exhibir una variedad de conductas diferentes.

esta característica biológica se apoya en observaciones etológicas sobre la conducta de animales y difiere del enfoque tradicional de la robótica que frecuentemente se concentra en el desarrollo de una sola conducta, a menudo excluyente de otras¹.

Preguntas robóticas que han sido “contestadas” por los seres vivos

Las tres preguntas anteriores han sido formuladas por [Varela88]

¿Puede la autonomía ser un resultado de formas de conductas de respuestas simples, dirigida por los datos del tipo estímulo/respuesta o es necesaria una conducta reactiva (dirigida por metas)?

Respuesta. En los seres vivos se encuentran ambas.

¿Es la memoria un requisito esencial para la autonomía?, si es así ¿cuál es la forma que debe tener esta memoria?

Respuesta. Los seres vivos cuentan con representaciones posiblemente icónicas.

¿Cómo coincidir el concepto de *auto-autonomía* (self-autonomy, la conducta que puede estar caracterizada como apoyo de la auto supervivencia) con el concepto de *autonomía impuesta* (imposed-autonomy, la conducta que no beneficia al robot o al ser vivo sino que satisface alguna tarea deseada que le imponemos)?

Respuesta. Coexisten ambas conductas.

¹ véase el trabajo sobre autonomía en sistemas deterministas autónomos de Covirgaru y Lindsay[1990]

Ninguna de estas preguntas pertenece al dominio de la psicología popular.

Modelos biológicos utilizados por la robótica.

Las investigaciones en robótica han comenzado históricamente con máquinas zoomórficas. Para ellas se han construido mecanismos, también zoomórficos para asir y manipular objetos y para que el robot se pueda desplazar o mover. Sin embargo, frecuentemente los controles han sido ajenos a los de los seres vivos.

Otros aspectos de la conducta de los animales han permitido el retomar el camino del control biológico, estos aspectos se relacionan con las conductas de los animales que les permiten interaccionar exitosamente en su medio ambiente. De lo anterior se ha derivado una metodología de construcción de robots denominada *basada en el comportamiento*, opuesta a la corriente tradicional en la que se diseñan robots mediante la instrumentación de una sola conducta global que frecuentemente es balística. La corriente nueva de la robótica pretende construir robots autónomos basados en conductas múltiples locales. Para poder derivar modelos de conductas de robots a partir de las conductas de los animales, es necesario entender cómo se generan tales conductas, cómo se interrelacionan y qué tipos de factores gobiernan la aparición de tales conductas.

La conducta animal: un enfoque roboticista

En nuestra opinión actualmente el diseño de sistemas basados en el comportamiento maneja una serie de conceptos biológicos comunes que se describen a continuación.

Las estrategias para estudiar la conducta animal se clasifican en dos grandes enfoques diametralmente opuestos, uno *descendente o analítico*, el de los etólogos y el otro *ascendente o sintético*, el de los neurofisiólogos como sus principales representantes. Bajo el primer enfoque se analiza y se explica la conducta sin hablar de los mecanismos *físicos* subyacentes. En el segundo enfoque se pretende explicar la conducta animal en términos de las

1 La conducta animal como referencia

interacciones entre los componentes físicos del animal, se intenta tener un conocimiento detallado de los diversos subcomponente físicos y comprender, a partir de ellos, los controles surgentes que expliquen la conducta.

Para los etólogos, la supervivencia de una especie dada, depende no sólo de su *estructura física* sino de contar con ciertas características conductuales para tal fin. Por ejemplo, no basta que un animal tenga las *características físicas* adecuadas para alimentarse sino que debe tener las conductas asociadas para procurarse el alimento.

Ya que no existe una teoría unificadora de las dos estrategias mencionadas y puesto que la descripción de la conducta es subjetiva y contextual, ya que depende de la conducta que se investiga, un método más completo de descripción es aquél en el que se describe la conducta en términos de sus consecuencias (actitud de diseño la llamaría Dennett²).

El modelo etológico clásico, propuesto por Lorenz, explica las conductas predecibles de ciertos animales ante ciertos estímulos, proponiendo que en el sistema nervioso central haya un mecanismo de liberación innato cuya función es causar una respuesta específica del animal cuando se detectan ciertas situaciones en el medio ambiente (véase la figura 1.1). El *liberador* es el estímulo que evoca tal respuesta y el patrón de conducta exhibido es llamado *patrón fijo de acción*. Este modelo, no explica completamente la aparición de ciertas conductas por lo que se ha recurrido a complementarlo, comenzando, para ello, por clasificar las conductas.

² Dennett, D. *La actitud Intencional*. Gedisa, 1990

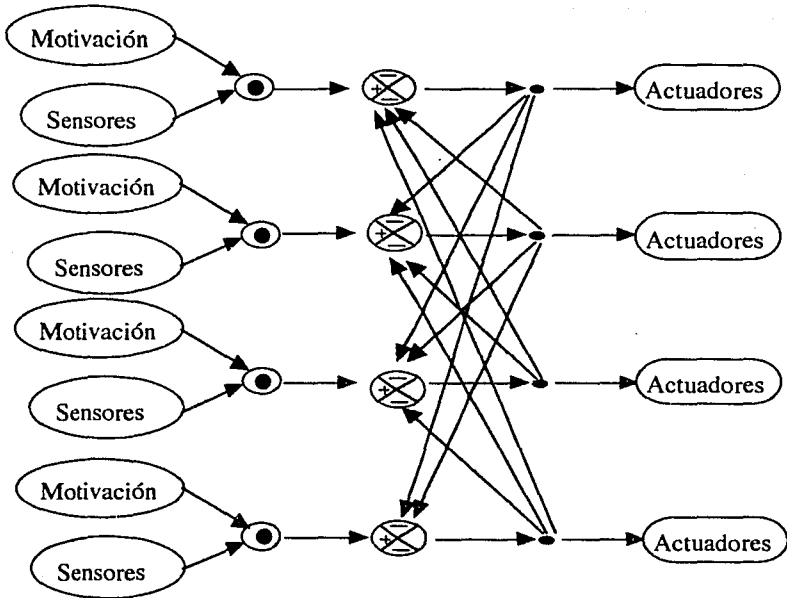


Figura 1.1 El modelo clásico de Lorenz, según [SchnepfU91]

Un concepto que los etólogos han utilizado para clasificar conductas se denomina *motivación*, es una manera de decir que ciertas actividades internas que apoyan ciertas conductas se pueden distinguir de aquéllas que soportan otras. Se dice que ciertas conductas están asociadas con las acciones de alimentación, caza, cortejo, etc. y que el animal tiene cierto potencial motivacional hacia la conducta exhibida.

Cada modo posible de conducta tiene asociado un potencial motivacional, de manera que el modo de comportamiento que se exhibe corresponde al de la

conducta que tiene mayor potencial motivacional y la combinación adecuada de estímulos externos. Por ejemplo, los potenciales motivacionales pueden ser una función del tiempo o de cambios hormonales. Estos potenciales motivacionales explican por qué algunas conductas se disparan o no. Se observan patrones complejos de conducta que frecuentemente se pueden descomponer *jerárquicamente* en términos de conductas más simples. Se dice que estas conductas complejas compiten por el control de conductas más simples o de nivel inferior. Se pueden presentar situaciones en las que ante un estado motivacional dado se den varias tendencias a realizar más de una actividad. Cuando surge la competencia entre estas tendencias se dice que ocurre una situación conflictiva que *requiere* un método para su solución. Se consideran dos formas de solución, una mediante un método de arbitraje entre tales conductas, y la otra a través de la activación de una conducta ajena. Resumiendo, la activación de conductas más complejas normalmente depende de restricciones internas y externas, es decir del estado interno y el del mundo exterior.

También se ha observado que los animales sólo responden a un pequeño subconjunto del total de información sensorial disponible en cualquier instante; de manera que los animales han desarrollado detectores especializados que les permiten identificar eventos específicos, por ejemplo, el sistema visual de la rana le permite detectar objetos que se mueven con una cierta velocidad y con una cierta forma: si en el ambiente en el que se encuentra hubiera posibles presas inmóviles la rana no las detectaría.

¿Por qué interesa observar la conducta de los animales?

Desde el punto de vista robótico se buscan los principios comunes más importantes que exhiben los animales en su existencia diaria, con la finalidad de encontrar cuáles de ellos pudieran ser útiles para el diseño y construcción de arquitecturas de robots. El estudio de tales conductas resulta interesante porque éstas tienen todos los ingredientes que se requieren para un agente autónomo artificial: orientación a metas, oportunismo, adaptatividad, plasticidad y robustez.

El principal ingrediente de todos ellos es la adaptatividad: un animal que confronta su ambiente ajusta su conducta continuamente.

Clases de conductas del etólogo

En etología las conductas se han clasificado atendiendo a: la naturaleza de sus reflejos componentes, la respuesta a los estímulos, la intensidad y duración de la respuesta, la intensidad y duración del estímulo e historia (memoria), entre las más importantes. Un ejemplo clasificatorio es el siguiente:

tactismos, tropismos y patrones de acción fija

Este grupo está formado por conductas que son consideradas *reflejas*. Hay otro grupo, el de las conductas *motivadas*, que aparecen no sólo por la presencia del estímulo ambiental apropiado sino que depende de lo que se denomina *variables motivacionales*. Estas conductas motivadas se caracterizan por:

- 1) Agrupar y serializar las conductas componentes en el tiempo.
- 2) Estar dirigidas por metas (objetivos): la secuencia de las conductas componentes sólo se puede entender por referencia a una meta;
- 3) Ser espontáneas: la conducta puede ocurrir en la ausencia total de cualquier estímulo evocador;
- 4) Ser cambiantes en la respuesta: el efecto *modulatorio* del estado motivacional varía, dependiendo del nivel de saciedad o estimulación;
- 5) Ser persistentes: la conducta puede durar más que el estímulo que lo inicia
- 6) Exhibir un aprendizaje asociativo.

La conducta de cualquier animal consiste en agregados de conductas *unitarias* (por ejemplo, tactismos, tropismos y patrones de acción fija) bajo algún control motivacional. Conforme el animal se enfrenta a su ambiente, debe coordinar apropiadamente muchas acciones posibles que resulten en conductas coherentes dirigidas hacia su supervivencia a largo plazo. El repertorio de conductas exhibe una organización, y la conmutación entre una conducta y otra depende de las condiciones ambientales y del estado interno. Las relaciones entre ellas no son sólo jerárquicas sino también heterárquicas. La organización depende del contexto conductual de forma tal que algunas

conductas se pueden sobreponer parcialmente (por lo que resultan difíciles de identificar); algunas tienen precedencia sobre otras; otras son mutuamente excluyentes, etc.

La plasticidad es otra característica que agrupa las conductas animales. Hay varias formas de plasticidad, pero en general se refieren a cambios de conducta futura imputables a la historia conductual pasada. La escala de tiempo de estas modificaciones tiene un *rango* que va desde segundos hasta años. Ejemplos de formas de plasticidad son:

habituación,	deshabitación,
sensibilización,	condicionamiento clásico,
condicionamiento instrumental,	aprendizaje latente

Las formas de plasticidad dependen del tiempo e intensidad del estímulo; de la aparición de otro estímulo; de eventos del ambiente, etc. Es a través de la plasticidad que los animales ajustan su repertorio de conductas (con el que están provistos genéticamente), a las exigencias del ambiente particular en el que se encuentran.

Los siguientes puntos resumen algunas observaciones importantes provenientes de la conducta animal que se han tomado en cuenta en la construcción de sistemas basados en el comportamiento .

- ✎ todos los animales poseen un conjunto de conductas innatas;
- ✎ el tipo de conducta exhibida en cualquier instante es el resultado de algún mecanismo interno de conmutación;
- ✎ una conducta compleja puede surgir como resultado de la aplicación secuencial de diferentes conjuntos de conductas primitivas unitarias. Las conductas primitivas actúan como mecanismos disparadores de la conducta siguiente;
- ✎ las conductas primitivas se organizan jerárquicamente.

Estudios etológicos y neuroanatómicos en batracios, pertinentes a nuestro trabajo

A continuación citaremos y comentaremos (en ocasiones con simples textos entre corchetes) lo que Rolando Lara Zavala escribió en 1987 [Lara87]:

...para poder comprender claramente los procesos que realiza el sistema nervioso para controlar la conducta animal, es necesario estudiar modelos biológicos que, por un lado, no sean tan complejos como los vertebrados superiores y por otro, tengan un sistema nervioso que nos permita extrapolar y comprender, tanto los procesos simples como [los] complicados de la función cerebral. Esta preparación, creemos nosotros, es la de los anfibios, en particular la de las ranas y sapos, ya que estos animales han sido estudiados extensamente desde un punto de vista etológico, fisiológico, anatómico, embriológico y teórico...

La situación *inversa*, que es la del espíritu de esta tesis, se podría expresar parafraseando a Lara de la manera siguiente:

para poder comprender claramente los procesos que realiza un robot reactivo para controlar su conducta, es necesario prestar atención a aquellos modelos biológicos que no sean tan complejos como los de los vertebrados superiores y que sin embargo, tengan un sistema nervioso que nos permita extrapolarlo a la robótica, ya sean los procesos simples como los complicados de la función cerebral. Esta preparación, creemos que es la de los batracios, en particular la de ranas y sapos en vista del extenso estudio biológico que se ha hecho de ellos.

Continuamos citando el trabajo de Lara:

...una gran cantidad de estudios etológicos nos ha permitido comprender lo que podíamos denominar la epistemología [la cognitividad] de los anfibios. Estos animales son sensibles a estímulos en movimiento, como a estímulos fijos, lo cual constituye la escena visual presente [se trata de una escena local], y son capaces de definir [emitir] la conducta motora que mejor adecue a esta situación específica, de acuerdo a la complejidad de su sistema nervioso central. Tal vez los estímulos más relevantes [los que en un momento dado obedecen a sus motivaciones] para los anfibios son los [desencadenados por objetos] que se mueven, tanto de tipo presa como depredador. Sin embargo, se ha observado que estos animales tienen una representación interna del espacio tridimensional [nosotros concebimos robots en espacios bidimensionales], con la cual pueden planear y ejecutar un gran número de acciones que les permiten cumplir adecuadamente [nosotros utilizamos también principios de adecuación robótica] sus necesidades primarias [motivaciones]:

alcanzar a la presa, huir del [de] predador, etc. Estos estudios han mostrado que los anfibios definen su conducta en base a la escena presente y al parecer, la ejecutan sin verificar continuamente si se acercan o se alejan de su objetivo: el sapo se acerca a la posición inicial de un estímulo en movimiento, sin importar su dirección, su velocidad, o su desaparición del campo visual ([tienen una conducta más *balística* que lo deseable para nuestros robots]. En un universo relativamente complejo, donde se encuentran barreras y zanjas, la conducta de los sapos depende de la relación de la posición de la presa y los obstáculos, y las dimensiones de estos obstáculos, lo cual sugiere que el animal tiene una representación métrica del espacio externo con medidas de profundidad [odometría] y dimensiones de los diferentes objetos [percepción de cualidades de los objetos], lo que le permite elegir adecuadamente la acción motora específica. En relación a este punto, se ha observado que los anfibios evalúan adecuadamente el tamaño de los estímulos tipo presa, siempre y cuando no estén alejados más de 30 centímetros.

En lo que respecta a estímulos en movimiento, los sapos son capaces de reconocer entre estímulo tipo presa [objetos-objetivo] y tipo depredador [objetos-peligrosos]: presa es un estímulo cuyo eje mayor se mueve en la dirección del movimiento, mientras que un estímulo de tipo depredador, es aquel que ocupa un área mayor de 20 grados o aquel cuyo eje mayor se mueve perpendicularmente en la dirección del movimiento

Por todo esto, podemos concluir que los anfibios consideran [nótese la terminología mentalista] que el objeto tiene carácter permanente sólo si se encuentra en su campo visual: tienen una representación práctica del espacio, es decir, no se representa el espacio como una estructura que les permite planear y evaluar diferentes conductas y sus consecuencias en forma simbólica [no cuentan con un modelo del *mundo*] sino que para cada situación específica se activa una conducta determinada [nótese el concepto implícito de unidad conductual]; no tienen un concepto de causalidad [ni siquiera un mecanismo predictor] ya que son incapaces de prever la posición de un estímulo en movimiento.

Estos animales, sin embargo, tienen cierta capacidad de representación, ya que pueden evaluar los cambios entre una escena pasada y una presente [memoria de largo alcance], específicamente en los fenómenos de habituación y aprendizaje...

...la retina de los sapos se proyecta en varias zonas del cerebro: el tectum..., el pretectum...[el talamo anterior]. Estas zonas a su vez, interaccionan entre sí [véase nuestra figura 1.2]:

1 La conducta animal como referencia

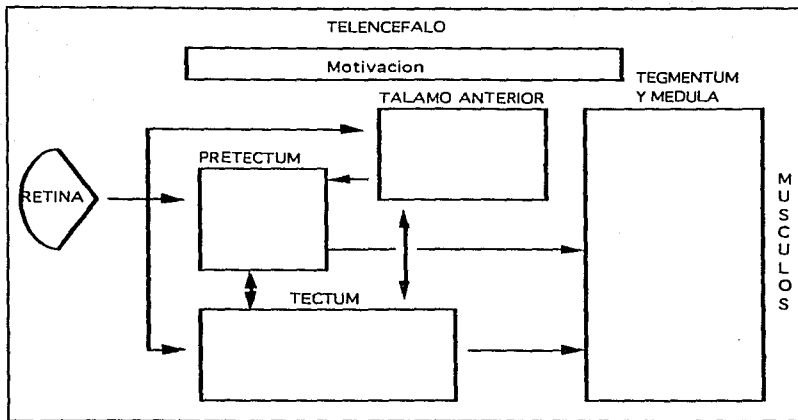


Figura 1.2. Nuestra interpretación del modelo neuro-anatómico de Lara del cerebro de la rana

El pretectum, tiene interacciones... con el telencéfalo; mientras que el tectum también establece una interacción con el telencéfalo a través del tálamo anterior... todo esto sugiere que la información en estos animales es procesada simultáneamente por varias regiones cerebrales que actúan intensamente entre sí para definir la respuesta motora del animal... a través del tegmentum, directamente a la médula espinal donde están las motoneuronas [véase la figura 1.2].

...el tectum parece jugar un papel relevante tanto en la respuesta de orientación como en la de huida, en combinación estrecha con el pretectum. Así mismo, estudios fisiológicos en esta zona han mostrado células que responden a la profundidad de los objetos.

El pretectum, por otro lado, es una zona que es sensible a estímulos de tipo depretratorio, aunque también, en combinación con el telencéfalo, tiene células relacionadas con procesos de memoria y aprendizaje. Esta zona aparentemente juega un papel relevante en la respuesta de huida y en el control de la respuesta de orientación. El tálamo anterior parece que, podría controlar la percepción de los objetos fijos, ya que al lesionarse esta zona el animal es incapaz de evadirlos y se han encontrado células que responden a este tipo de estímulos. Esta zona también es sensible al color del medio ambiente y juega un papel relevante en el control de la

1 La conducta animal como referencia

respuesta fototáctica, es decir, la tendencia a buscar los lugares luminosos de color azul.

Finalmente, al telencéfalo se le han asignado funciones relacionadas con el estado motivacional del animal [nótese otra vez la terminología mentalista], ya que el hipocampo es la región que controla la mayor parte de la regulación de las variables internas del animal, y con procesos de memoria y aprendizaje, ya que al lesionarse esta zona, modifica la capacidades de habituación del animal...

El trabajo de entender el cerebro por parte de los investigadores de la Inteligencia Artificial parece ocurrir como si se tratara de una máquina eléctrica con entradas y salidas de sensores y actuadores del cuerpo. Según Brooks los sistemas biológicos reales no son agentes racionales que toman datos de entrada, computan lógicamente y producen salidas. Son una mezcla de muchos mecanismos que trabajan de varias maneras de las cuales surge el comportamiento que observamos como racional. Esto se puede observar con más detalle cuando comparamos los individuos a nivel computacional y a nivel de organización del cerebro.

Dado que no entendemos totalmente ni aún a los animales más simples, no se puede esperar una comprensión suficiente para construir inteligencia artificial solo por la mera observación de estos sistemas nerviosos y menos de la de los sistemas nerviosos centrales de animales más complejos. [Cruse91] y otros han proporcionado ciertas ideas de cómo trabajan los animales simples basándose en la comprensión de niveles primitivos de sus circuitos neurológicos. Este tipo de claves pueden ayudarnos para construir robots que caminan como es el caso de la *neuroetología computacional* de [Beer-Chiel91] y del trabajo de Brooks. Todo ello puede ayudarnos a construir mejores sistemas artificiales aunque por sí mismos no puedan proporcionarnos una teoría completa.

2 La cognitividad inherente a la instrumentación en Inteligencia Artificial

El presente capítulo inicia la discusión de cognitividad como una propiedad de nuestra intuición del término a través de los *ojos* de una tecnología: la Inteligencia Artificial clásica. Lo anterior permite definir cognitividad en términos *físicos* tomando como elementos de discurso el símbolo físico, las formas de representación y las manipulaciones de ésta.

A diferencia de lo anterior, los sistemas enactivos de Varela apuntan en la dirección de nuestro proyecto de tesis en el sentido de la surgencia de un control, sin símbolos físicos ni representaciones.



2La cognitividad inherente a la instrumentación en Inteligencia Artificial

El nombre de Ciencia Cognitiva designa a las disciplinas involucradas en *el análisis científico moderno del conocimiento en todas sus dimensiones* [Varela88]. A partir de la fase moderna del cognitivismo *el conocimiento se ha ligado tangiblemente a una tecnología que transforma las prácticas sociales que lo posibilitaron..* La inteligencia artificial es un ejemplo visible de esta liga. Debido a que uno de los papeles de la tecnología es el de actuar como un amplificador de estas ideas no es posible separar las ciencias cognitivas de la tecnología cognitiva puesto que se perdería un elemento que resulta vital para entenderlas, además de que existe una realimentación de las ideas entre ambos campos. La inteligencia artificial, según Varela es considerado como el polo tecnológico de la ciencia cognitiva del cual se derivan tecnologías tales como: el reconocimiento de imágenes, la comprensión del lenguaje, la síntesis de programas y la robótica, entre otras.

Hay una realimentación de los resultados obtenidos en las diversas disciplinas de manera tal que se reconoce que el cognitivismo moderno ha pasado por varias etapas cada una de ellas asociada o derivada de técnicas de la inteligencia artificial.

La primera etapa de la fase moderna de la Ciencia Cognitiva comenzó con el impacto de la cibernética que dió lugar a *un movimiento cuya intención era crear una ciencia de la mente* [Varela88], en esta fase, la pretensión era la de expresar los procesos subyacentes de los fenómenos mentales en mecanismos explicativos y formalismos matemáticos, esto a su vez dió lugar a: uno proponer a la lógica como la disciplina adecuada para comprender el cerebro y la actividad mental, y dos a ver el cerebro como un dispositivo que encarna principios lógicos en sus elementos constitutivos, de manera tal que el cerebro entero se podía considerar como una máquina deductiva. Bajo este enfoque cognitivista, la conducta inteligentes supone la capacidad para representar el mundo de ciertas maneras. De tal modo que la conducta cognitiva se explica dando por sentado que un agente actúa representando rasgos relevantes de las situaciones en las que se halla, y conforme su representación de la situación sea exacta, la conducta del agente tendrá éxito. En este enfoque se parte de la hipótesis de que la cognición consiste en actuar sobre la base de *representaciones que adquieren realidad física en la forma de un código simbólico en el cerebro o en la máquina.* Se considera que esta

2La cognitividad inherente a la instrumentación en Inteligencia Artificial

computación es fundamentalmente semántica o representacional. *No hay computación sin representación.*

Bajo este enfoque se reconocen tres niveles para explicar la cognición: el físico, el simbólico y el representacional o semántico. De manera que una expresión simbólica poseedora de una realidad física se puede corresponder con un patrón global altamente distribuido de actividad cerebral.

Esta multiplicidad de niveles de la explicación científica constituye una de las principales innovaciones de la Inteligencia Artificial.

Sin embargo, hay en la literatura muchas inconformidades respecto a los puntos de vista establecidos por este enfoque de la Inteligencia Artificial, que se materializan en las dos siguientes formas:

- La crítica de la computación simbólica en cuanto portadora adecuada de las representaciones.
- La crítica de la pertinencia de la noción de representación en cuanto componente fundamental de la Inteligencia Artificial.

Una alternativa a la de la orientación simbólica es la *surgencia* de conocimiento en la estrategia conexionista. Esta estrategia resuelve dos problemas presentes en el cognitivismo simbólico: uno es la limitación impuesta por el procesamiento de información simbólica basado en reglas *secuenciales*, aplicadas una por una. Lo que se convierte en una fuerte limitación cuando la tarea a ejecutar requiere gran cantidad de operaciones secuenciales (considérese el caso del análisis de imágenes o pronóstico del tiempo). El segundo problema importante es que el procesamiento simbólico está localizado. Lo cual da lugar a que cualquier disfunción o pérdida de cualquier parte de las reglas o de los símbolos produzca un daño grave. Por lo tanto es muy deseable una operación distribuida para que halla una equipotencialidad e inmunidad ante las mutilaciones o fallas.

La estrategia conexionista consiste en construir un sistema cognitivo a partir de componentes simples que se conectarían dinámicamente entre sí de manera

2La cognitividad inherente a la instrumentación en Inteligencia Artificial

densa, por ejemplo el caso de las redes neuronales, o el de los autómatas celulares. Bajo este enfoque cada componente opera sólo en su ámbito local, sin un agente externo que los supervise o dirija. Hay una cooperación *global* que *surge* espontáneamente cuando todos los componentes participantes alcanzan un estado mutuamente satisfactorio. Sin embargo, no hay una teoría unificada de las propiedades surgentes y se habla de ellas como propiedades globales, dinámica de la red, redes no lineales sistemas complejos, o de sistema sinérgicos.

Así el cerebro como fuente de metáforas e ideas para otros campos se puede ver no a partir de descripciones simbólicas abstractas, sino como un conjunto de componentes de tipo neural, no inteligentes, que apropiadamente conectados presentan interesantes propiedades globales, tales como la autoorganización. Esta se hace corresponder con las aptitudes cognitivas que se estudian. Se observa que el enfoque depende de las conexiones apropiadas, lo cual se realiza mediante reglas de aprendizaje.

Las teorías conexionistas brindan modelos funcionales para diversas e interesantes aptitudes cognitivas, tales como el reconocimiento rápido, la memoria asociativa y la generalización categórica. Los modelos conexionistas están mucho más cerca de los sistemas biológicos, así que se puede aspirar a trabajar con un grado mayor de integración entre la inteligencia artificial y la neurociencia.

Uno de los aspectos más interesantes de este enfoque es que los símbolos no desempeñan ningún papel. La computación simbólica es reemplazada por operaciones numéricas. Bajo el modelo conexionista el resultado de una sola computación simbólica discreta se obtendría a través de una gran cantidad de operaciones numéricas que gobiernan una red de unidades simples. Los componentes significativos no son símbolos sino patrones complejos de actividad entre las muchas unidades que forman la red. De esta manera se dice que el significado (o sentido) no está localizado en símbolos particulares sino que éste es función del estado global del sistema.

2La cognitividad inherente a la instrumentación en Inteligencia Artificial

La enacción una alternativa frente al conexionismo.

El problema en la definición de cognición de los enfoques anteriores se basa en la total ausencia del sentido común. Para esos enfoques el criterio principal de cognición es la representación atinada de un mundo externo que está dada de antemano. Sin embargo, actualmente se considera que la mayor capacidad de la cognición viviente consiste en gran medida en *plantear* las cuestiones relevantes que van surgiendo en cada momento de nuestra vida (no son predefinidas sino que se les hace emerger desde un trasfondo) y lo relevante son aquéllas que nuestro sentido común juzga como tal, siempre dentro de un contexto. Así el problema mencionado anteriormente trata de una crítica hacia la noción de representación como núcleo de la Inteligencia Artificial, dejando a la búsqueda del *paralelismo distribuido* como una solución de tipo pragmático. Esta crítica se basa en que en el mundo en que vivimos va surgiendo o es modelado en vez de estar predefinido. Hay un interés central en el fenómeno de la *interpretación* entendida como la actividad circular que eslabona la acción con el conocimiento, al conocedor con lo conocido, todo en un círculo indisoluble. Esta perspectiva analítica enfatiza la acción más que la representación por lo que Varela la denomina *enacción*

En el campo de la Inteligencia Artificial esto se refleja de la manera siguiente: se ha supuesto que es posible dividir el mundo en dominio. Por ejemplo, en el juego de ajedrez es fácil definir su forma, el espacio del juego de ajedrez consiste de estados posibles. Hay piezas y posiciones en el tablero. Hay reglas de movimientos y turnos. Hay límites definidos claramente. Sin embargo, el enfoque usado para resolver el juego de ajedrez resulta infructuoso para el dominio de los robots móviles. El dominio en el que se debe desempeñar un robot móvil no está definido claramente.

La categorización de cualquier aspecto del mundo natural no tiene límites precisos; no se puede expresar como un dominio a partir del cual elaboramos un mapa.

Para Varela, en este enfoque la operación del cerebro se centra en la enactuación de mundos a través de la historia de las generaciones viables: *el*

2La cognitividad inherente a la instrumentación en Inteligencia Artificial

cerebro es un órgano que construye mundos en vez de reflejarlos.

Para dar una síntesis de las diferencias entre las etapas de la ciencia cognitiva y de la inteligencia artificial como su representante tecnológico, Varela propone tres preguntas básicas que deben ser contestadas por cada enfoque, ellas son:

1. ¿qué es la cognición?
2. ¿cómo funciona?
3. ¿cómo saber si un sistema cognitivo funciona adecuadamente?

Respuestas de la cognitividad simbólica

1. Procesamiento de información: manipulación de símbolos basada en reglas.
2. A través de cualquier dispositivo que pueda representar y manipular elementos físicos discretos: los símbolos. El sistema interactúa sólo con la forma de los símbolos (sus atributos), no su significado.
3. Cuando los símbolos representan apropiadamente un aspecto del mundo real, y el procesamiento de la información conduce a una feliz solución del problema planteado al sistema.

Respuestas conexionistas

1. La emergencia de estados globales en una red de componentes simples.
2. A través de reglas locales que gobiernan las operaciones individuales y de reglas de cambio que gobiernan la conexión entre los elementos.
3. Cuando vemos que las propiedades emergentes (y la estructura resultante) se corresponden con una aptitud cognitiva específica:

2La cognitivdad inherente a la instrumentación en Inteligencia Artificial

una solución feliz para la tarea requerida.

Respuestas enactivas

1. Acción efectiva: historia del acomodamiento estructural que enactúa (hace surgir) un mundo.
2. A través de una red de elementos [virtualmente] interconectados capaces de cambios estructurales durante una historia ininterrumpida.
3. Cuando se transforma en parte de un mundo de significación preexistente [se empotra (embed) como lo hacen los vástagos de toda especie], o configura uno nuevo [como ocurre en la historia de la evolución].

Para este enfoque, las representaciones ya no desempeñan un papel central, la *inteligencia* ha dejado de ser la capacidad de resolver un problema para ser la capacidad de ingresar en un mundo compartido

La construcción de agentes o robots explícitamente cognitivos es el tema del presente capítulo. La inteligencia artificial distribuida ha encontrado que, entre la posible gama de agentes para lograr una inteligencia distribuida se encuentran aquéllos a los que se les asignan representaciones mentalistas explícitas manipuladas dentro de una racionalidad explicitada.

Los *modelos mentales* del agente están generalmente manipulados por operadores modales de creencias, compromisos y capacidades.

Dada la explicitéz de estos modelos, no son un campo adecuado para encontrar los principios que buscamos porque prácticamente están dados en esta construcción.



Se ha dicho que la Inteligencia Artificial clásica se circunscribe a “la construcción de programas informáticos capaces de ejecutar tareas complejas, con base en una centralización y concentración de la *inteligencia*” en el seno de sistemas únicos”. Sin embargo, también se reconoce que esto genera un gran número de dificultades debidas en su mayor parte, a la necesidad de integrar en el seno de una misma base de conocimientos, la pericia, las competencias y los conocimientos de individuos diferentes para la realización de un fin común.

El enfoque distribuido de la Inteligencia Artificial, que suele denominarse Inteligencia Artificial Distribuida, propone distribuir la inteligencia entre numerosos agentes, no sujetos a un control centralizado. Los sistemas así generados están constituidos por un grupo de agentes, cada uno dotado de una cierta autonomía, pero que en conjunto son capaces de planificar, actuar y trabajar en un medio común, bajo la presión de demandas eventuales. Estos sistemas ya no son más, *pensadores encerrados en sus propios pensamientos*, sino pensadores abiertos hacia el exterior, capaces de hacer cambios, en el seno de un grupo o en el seno de verdaderas sociedades. De aquí la aparición de nociones nuevas en Inteligencia Artificial tales como: cooperación, coordinación de acciones, negociación, etc.

Puesto que es difícil modelar la experiencia global de numerosos elementos, se han diseñado técnicas que permiten formalizar el comportamiento de un conjunto de individuos que interactúan para la realización de su objetivo.

El consiste frecuentemente en resolver la circunstancia de una *distribución material* comp puede ser la de diferentes centros que constituyen una red. Ahí donde el abordaje clásico busca representar todo, en un modelo único de funcionamiento de conjunto, cada centro (como entidad física) es considerado como un agente autónomo, interactuando con otros agentes. Ejemplos de lo anterior son una red heterogénea de computadoras o de transportes o una red de distribución de energía. Estas redes están constituidas por nodos centrales, cada uno de los cuales puede comprender varios sistemas expertos destinados al análisis de fallas y de riesgos, por ejemplo. El objetivo de un sistema como éste es el de planificar el mejor funcionamiento del conjunto y su

explotación, respetando la arquitectura y los materiales existentes.

Otro problema puede ser el de la *distribución funcional*, la experticia puede, en contraste con los ejemplos anteriores, ser abordada en términos de funciones verdaderas. Cada función verdadera requiere, la mayor parte del tiempo, la cooperación de numerosos expertos. Esta situación se encuentra, por ejemplo, en la conducción de naves tales como los supertanques, en la conducción de los portaviones, en la planificación de misiones espaciales, etc. Aunque los peritos se apoyan sobre informaciones idénticas y la conceptualización conjunta del problema es susceptible de conducir a conclusiones divergentes según los criterios de apreciación, expresados éstos bajo la forma de las reglas. Aquí el acercamiento mediante la Inteligencia Artificial Distribuida, consiste en elaborar una arquitectura constituida de entidades expertas independientes, esperando poder resolver situaciones conflictivas que surjan entre las entidades.

Los problemas de planificación se solucionan en la Inteligencia Artificial no-distribuida combinando, a veces, la investigación de operaciones con la Inteligencia Artificial no-distribuida, pero esto obliga muy a menudo a detener el conjunto del proceso y a reiniciarlo, para poner en juego un nuevo plan. La Inteligencia Artificial distribuida por el contrario, ofrece las posibilidades de negociación, permitiendo soluciones locales de conflictos y una replanificación dinámica.

La Inteligencia Artificial Distribuida conduce también a la construcción de sistemas que permiten movilizar el comportamiento de un conjunto de entidades más o menos expertas, más o menos organizadas, respetando *leyes de tipo social* de estas entidades o agentes. Estos sistemas disponen de una cierta autonomía y están inmersos en un conjunto en el cual y con el cual interaccionan. De aquí, la división de su estructura individual en tres funciones principales: la de percibir, la de decidir, y la de actuar. Los agentes disponen, para hacer esto, de una representación parcial de su medio y de medios de percepción y de comunicación. Su comportamiento está fundamentado sobre conocimientos dados *a priori* y sobre conocimientos adquiridos.

Los agentes, de una manera, general ejecutan dos grandes funciones: *una de tendencia social*, externada hacia la colectividad (está por lo tanto en el dominio de lo conversacional (lingüístico) y de lo relacional, son mecanismos, y bases de conocimiento asociadas, concernientes a actividades de grupo. La otra de *tendencia individual* (conciérne al Saber y a lo conductual, donde los mecanismos y las bases de conocimiento contienen las reglas de funcionamiento interno del agente).

Está claro que un sistema multiagente (agente con funciones individuales diferentes) debe, por lo tanto, integrar ciertos mecanismos que aseguren la dinámica de participación de un individuo en la vida de un grupo. Los principios son bien conocidos por el especialista de la dinámica de grupo y por los psicólogos: forman un entorno de disciplinas que hacen intervenir las ciencias de la computación pero también la *sociología*, la *psicología* y la *etología*.

En cuanto al nivel de inteligencia que caracteriza a cada agente, ésta varía según dos grandes tipos de opciones en el dominio de la Inteligencia Artificial Distribuida. Una de alto nivel que corresponde a la escuela que nosotros llamamos cognitiva explícita (mentalista explícita) fundada sobre la cooperación de agentes ya capaces por ellos mismos de realizar operaciones relativamente complejas. Otra, la del bajo nivel (no mentalista o cognitiva implícita), que se desarrolla rápidamente en nuestros días, es la de agentes reactivos (véase el capítulo 2) donde los agentes son más simples, menos inteligentes, pero pueden ser mucho más numerosos, las diferencias entre estos dos enfoques se discutirán en éste y en los capítulos siguientes.

Los agentes "cognitivos explícitos"

En la escuela cognitiva explícita la distinción de conocimientos en numerosas entidades, obliga a concebir a las bases de conocimientos, no ya como únicas y verdaderas en sí, sino ligadas a realidades del mundo, es decir; que el agente puede saber de otras bases por una parte y puede conocer su medio por otra. El tomar en cuenta esta exterioridad, compuesta a la vez de consecuencias de actos de cada agente y comunicaciones interagentes,

requiere del desarrollo de teorías en el dominio de protocolos de comunicación, en el de cooperación de agentes y en el de su relación con el medio ambiente.

La coordinación de las acciones es uno de los problemas mayores que deben afrontarse dentro de los dos esquemas principales: ya sea que se trate de disponer de un sistema central capaz de determinar y de planificar globalmente las acciones de los diferentes agentes, o en sentido inverso, se decida dar una total autonomía a los agentes, que por lo tanto, deben identificar ellos mismos los conflictos para resolverlos localmente. La elección de uno u otro camino, juega a la vez, un papel definitivo sobre la arquitectura de cada agente y sobre las estructuras de comunicación.

Por otro lado, dentro de una dinámica colectiva de las acciones de cooperación, se exigirá que el agente posea un cierto número de cualidades concernientes al grado y la naturaleza de su inserción en esta dinámica, es decir, de sus *compromisos*, de sus *finés individuales*, de sus *intenciones*, de sus *creencias* y del *modelo que tiene de sí mismo, de otros, y de su entorno* o ambiente (en esto estriba su explicitéz). Estos requisitos implican para nosotros la imposición de principios lo que excluye su descubrimiento como es nuestro objetivo.

La noción de compromiso coincide con la acepción habitual que éste término tiene y constituye una de las cualidades esenciales de los agentes cooperativos. Un agente cuando promete hacer una acción, se compromete a su vez con otro agente, lo que lo conduce a actuar en consecuencia. Pero la noción de compromiso, hace igualmente referencia a otro tipo de conducta, la que se podría calificar de compromiso con el *mundo*. Esta forma de compromiso traduce la idea de que toda acción de un agente modifica de manera irreversible el estado del sistema en el cual se desempeña.

Así, la acción de un agente se encuentra condicionada por esta irreversibilidad que se debe tomar en cuenta, a diferencia de las técnicas de Inteligencia Artificial clásica que suponen que en todo instante hay un regreso posible al estado anterior. El tiempo representa, en este aspecto, un

parámetro fundamental en el desarrollo y encadenamiento de las acciones efectuadas. Esto no implica que una misma temporalidad sea compartida por todos los agentes, que son en principio autónomos y asíncronos. Para administrar el paralelismo de ciertas tareas, o para sincronizar u ordenar las acciones, el final de la ejecución de cada operación puede servir de señal para la *puesta en marcha* de las operaciones siguientes. Cada agente dispone por lo tanto de su propio reloj. Sin embargo, se pueden concebir sistemas multiagentes sin ningún reloj de referencia global.

En cuanto a la intención de un agente, se define ésta como la voluntad consciente de lograr un fin o de efectuar una acción. La acción puede ser *inmediata*, es decir, intrínsecamente ligada a la acción subyacente y al fin que persigue.

Es posible organizar localmente los discursos de los agentes, definiendo *estructuras de diálogo* y cambios que se apoyan únicamente sobre las acciones que determinan el papel que debe jugar un agente en un momento dado. Cuando nosotros hablamos, realizamos verdaderos actos que modifican las creencias de nuestros interlocutores, en el sentido de una sentencia es relativo a una acción que, como todas las acciones puede tener éxito o fracasar. Así toda acción de un agente está ligada a un estado intencional que comprende sus fines, sus voluntades, etc.

Los diálogos se descomponen en actos de lenguaje [Searle80], lo que da lugar a que las estructuras de comunicación se vuelvan homogéneas, son estructuras del tipo *percepción/acción*.

Un agente puede igualmente tener *intenciones futuras*, es decir, proyectos de realización de acciones más o menos diferidas, formando parte integral de un plan.

Como quiera que se definan los fines individuales, la toma de compromisos o el anuncio de intenciones: todos estos componentes, suponen, por encima de todo mecanismos de razonamiento apropiados (agente racional) y una *lógica subyacente*. Los agentes deben tener en cierta medida, la posibilidad de

adquirir informaciones sobre el medio y sobre los otros, sea directamente por su propio sistema de percepción, sea por intercambios de información, sea más aún, por un aprendizaje. Las informaciones deben en consecuencia, encontrar una representación formal que les permita ser utilizables por los mecanismos de razonamiento, de manejo de hipótesis, de toma de decisiones, etc.

Además de su limitada autonomía y de sus capacidades limitadas, un agente no tiene, del medio que lo rodea y de los otros, más que una representación parcial a base de modelos incompletos.

Debido a lo anterior el agente, en su participación en la dinámica colectiva, debe tomar decisiones que pueden ser irrealizables o que no satisfagan los objetivos visualizados. La solución de esos problemas se encuentra en el uso de *razonamientos revisables*, en *métodos de validación de hipótesis*, y en la *utilización de lógicas no clásicas*, tales como lógicas borrosas, lógicas modales o lógicas de la incertidumbre.

Las relaciones entre el saber del agente y el mundo, supone el establecer por parte del agente una correspondencia entre el mundo y su representación. Esta correspondencia es una interpretación, es decir, una función que se explica sea en términos de verdad, o en términos de funciones de la lógica, o en términos de entidades del mundo, a las cuales esta representación hace referencia. Esta, debe permitir determinar lo que es la aceptación de una información: el *como-saber* implica la semántica del mundo imaginado.

Una vez provistos de estas funcionalidades, los agentes deben, contar con el *poder de comunicarse*, a fin de actuar: comunicarse sus interacciones para sincronizar sus acciones y para poder resolver conflictos, bien sean conflictos de recursos, de fines o de intereses. Sus comunicaciones, igualmente, se hacen para ayudarse mutuamente o para suplir los límites de su campo de percepción. Un agente, en efecto, no puede estar en todo momento en relación con todos los otros, ni estar equipado con todos los captadores necesarios para el conocimiento del medio.

Existen actualmente dos tipos de arquitectura para los sistemas de comunicación. Una consiste en instalar un procedimiento de comunicación compartida. Conciérne esencialmente a agentes evolucionados, este procedimiento utiliza un espacio común, *un pizarrón* (blackboard), dentro del cual el agente escribe o lee las informaciones. Una segunda opción, utilizada indistintamente por los agentes cognitivos explícitos o por los reactivos, se reduce a intercambios asíncronos y discontinuos de mensajes mínimos de un agente a otro. Un protocolo permite estructurar y asegurar la continuidad de las comunicaciones. Cada agente emite mensajes, comenzando por una palabra clave de principio y otra de terminación. Estos mensajes son órdenes o informaciones, requisiciones o preguntas, promesas o amenazas, etc. Las respuestas son los mensajes portadores de información o las acciones. *Los protocolos son tomados de principios psicocognitivos y sociales [están explicitados].*

Los sistemas de comunicación son, de hecho, más o menos elaborados, su complejidad está directamente ligada a las capacidades cognitivas y a la arquitectura interna de los agentes.

En la figura 3.1 se muestra la clasificación de agentes propuesta por [Erceau -Ferber91]. En ella se aprecia la ubicación de los agentes reactivos en el nivel más bajo y los agentes intencionales en el extremo opuesto, esta última clase de agente posee todas las características que hemos discutido en párrafos anteriores.

Para [Shoham90], los agentes (los robots entre ellos) son entidades que poseen versiones formales de un *estado mental*, cuyo contenido son las creencias, compromisos, capacidades y decisiones del agente.

Para McCarthy, lo importante de estas descripciones mentalistas es el poder que tienen para acceder al diseño robótico: auxilian a entender el funcionamiento pasado o futuro de los robots, y también ayudan a comprender la estructura del robot ya sea para repararlo o para mejorarlo.

Es posible pasar de una descripción mentalista a la descripción de procesos

de bajo nivel (instrumental). Esta traducción es el proceso denominado *agentificación* por [Shoham89]. Es necesario contar con un medio, proceso o herramienta que dada una representación conductual del proceso o descripción mecanicista, la transforme en un conjunto de programas de nivel intencional, o viceversa.

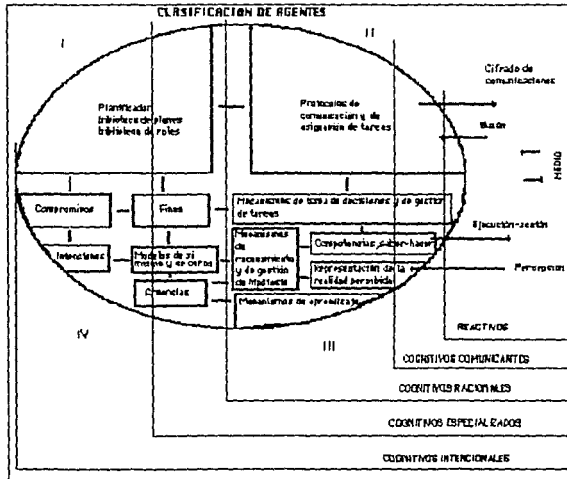


Figura 3.1 Los cuadrantes superiores representan al SER SOCIAL, sus relaciones con su medio y con otros, y el universo con el que se comunican. El cuadrante I concierne a la dinámica de sus relaciones y de su actualización, mientras que el cuadrante II contiene los mecanismos necesarios para su elaboración y planificación. El manejo y el tratamiento de mensajes recibidos en el BUZÓN permite al agente saber qué hacer, dónde se encuentran los otros agentes, sus planes, etc. Los dos cuadrantes inferiores representan al ser individual del agente, es decir sus características propias. El cuadrante III contiene las capacidades PERSONALES de percepción, de decisiones y de acción sobre el MUNDO, así como los mecanismos que se le asocian. Estos mecanismos determinan el comportamiento del agente. El cuadrante IV comprende los modelos que el agente tiene de sí mismo, de otros y del UNIVERSO. Las líneas de demarcación verticales permiten ver la complejidad creciente (de derecha a izquierda) de los agentes desde los agentes reactivos hasta los cognitivos más elaborados, los que contienen todas las funcionalidades.

4 Diseño cognitivista implícito

En este capítulo
revisamos la "filosofía"
de un nuevo paradigma
de la Inteligencia
Artificial, el de los
*sistemas basados en el
comportamiento*,
fundamento de nuestra
experimentación



El capítulo 1, en el que hablamos de los aspectos etológicos, psicológicos y neurofisiológicos de los animales, obviamente no comprende todo lo pertinente a nuestra tesis, pero es suficiente para mostrar que en los sistemas biológicos, la inteligencia está organizada de formas muy diferentes de los puntos de vista tradicionales de la Inteligencia Artificial. En este capítulo, se presentan las ideas de SITUACIONALIDAD e INCORPORACION que son dos piedras miliars de la nueva manera biológica de ver la Inteligencia Artificial, y se discute la INTELIGENCIA, y su surgencia bajo estos contextos. Estas ideas claves han conducido a un nuevo estilo de investigación en Inteligencia Artificial que es denominado sistemas basados en el comportamiento.

La situacionalidad

La Inteligencia Artificial tradicional adoptó un estilo de investigación donde los agentes que se construyen, para probar teorías en inteligencia, son esencialmente solucionadores de problemas, que trabajan en un dominio simbólico abstracto. Los símbolos pueden tener referentes en las mentes de los constructores de los sistemas pero no hay nada que *aterrice* esos símbolos al mundo real, los agentes no están situados en un mundo.

En estos sistemas no hay un mundo externo *per se*, continuo y contingente. Los sistemas tratan únicamente con un modelo interno del mundo, con su propia física interconstruida, veáse el capítulo 5. Es común que en la mente del programador no esté claro la diferencia entre el conocimiento del agente y el mundo en el que se supone debe operar --muchos sistemas de Inteligencia Artificial se eligen para que no haya distinción entre el mundo y el conocimiento: el agente tiene acceso a una percepción directa y perfecta, y una acción directa y perfecta. Los modelos de la representación del conocimiento que han sido desarrollados en Inteligencia Artificial clásica esperan y requieren de una correspondencia uno-a-uno entre el mundo y la representación que de él tiene el agente.

Los primeros robots que se construyeron seguían este enfoque: construían modelos del mundo, planeaban trayectorias alrededor de obstáculos y

actualizaban sus estimaciones de la ubicación de los objetos relativos a ellos cuando se movían, véase el capítulo siguiente sobre los antecedentes de la robótica no-reactiva.

Brooks ha desarrollado un enfoque diferente, ha construido robots móviles que usan el mundo como su propio modelo. Es decir, cada robot continuamente activa sus sensores en lugar de consultar un modelo interno del mundo. Las ventajas de este enfoque es que los problemas asociados con la clasificación objetos e identidades desaparecen. Un resultado inmediato de estas ventajas es que el procesamiento perceptual se vuelve mucho más simple y el desempeño del robot para realizar tareas es mejor y con mucho menor cómputo, aunque para esto se requiere contar con una mayor diversidad de modalidades *sensoriales*.

Un robot debe responder a sus entradas en tiempo real y si utiliza un modelo del mundo la computación que requiere para consultarlo va en detrimento del tiempo de su repuesta. La continuidad que se necesita la puede proporcionar el mundo mismo, de manera que resulta más eficiente la percepción del mundo que la consulta un modelo del mundo.

La incorporación

El aterrizaje físico de un robot en el mundo fuerza a su diseñador a contender con todos los aspectos del mismo. Si el agente inteligente tiene un cuerpo, tiene sensores y tiene actuadores, entonces todos los detalles y aspectos de estar en el mundo se deben encarar. Ya no es posible arguir que el sistema perceptual simulado es realista, o que el problema de la incertidumbre en la acción no es significativo. Cuando se hacen experimentos físicos repetidamente, es común encontrar que muchos de los problemas que parecían significativos no lo son en el sistema físico, y muchos de los que no parecían problemas se vuelven más grandes.

Hay dos razones por las que la incorporación de sistemas inteligentes es crítica. Primero, solamente un agente inteligente incorporado está totalmente validado en tanto que puede entender el mundo real. Segundo, únicamente a

través de un previo aterrizaje físico cualquier símbolo interno u otro sistema puede encontrar un lugar para cambiar de nivel y dar *significado* al procesamiento que se lleva a cabo dentro del sistema.

Un problema más profundo es: *¿puede haber una mente desincorporada?*. Muchos creen que lo que es humano en nosotros está relacionado muy directamente con nuestras experiencias físicas. Por ejemplo [Johnson87], arguye que una gran cantidad de nuestro lenguaje está relacionado metafóricamente a nuestras conexiones físicas en el mundo. Nuestros *conceptos* mentales están basados en ejemplares experimentados físicamente. [Smith80] sugiere que sin *aterrizaje* físico no puede haber un alto en la *regresión filosófica* dentro de los sistemas basados en el conocimiento: tal como tratar de razonar acerca del conocimiento del mundo real en la forma como está contenido en una enciclopedia.

Sin una participación y percepción del mundo activos no hay significado para un agente. Todo resulta ser símbolos aleatorios. Los argumentos que se podrían hacer en contra son que en un nivel cualquiera de abstracción, aún la mente humana opera en esta posición solipscista. Sin embargo, la evidencia biológica (véase el capítulo 1) sugiere que la conexión de la mente con el mundo es tan fuerte, y con tantas facetas, que las abstracciones filosóficas solipscista pueden no tener sentido.

La inteligencia

[Brooks91] argumenta que las clases de actividades que pensamos como típicas que demuestran inteligencia en las personas han ido tomando lugar en una fracción muy pequeña de nuestro linaje evolutivo. Aún más, arguye que las cosas *simples* que se realizan con percepción y movilidad en un ambiente dinámico le tomó a la evolución mucho tiempo para perfeccionarlas y que esas capacidades son una base necesaria para el intelecto a *nivel más alto*. Por lo tanto, propone observar a animales más simples como modelo ascendente para construir inteligencia. Si se considera que el *razonamiento* es el componente principal del intelecto del robot, entonces la dinámica de las interacciones del robot con su medio ambiente son determinantes primarios

de la estructura de su inteligencia.

Antes que Brooks, H. Simon había discutido un punto similar en términos de una hormiga que camina sobre la playa. Puntualizaba que la complejidad de la conducta de la hormiga es más un reflejo de la complejidad de su medio ambiente que de su complejidad interna. Pero, desgraciadamente redujo el estudio de la conducta humana al dominio de los problemas cripto-aritméticos.

Es difícil dibujar la línea de que lo que es la inteligencia y lo que es la interacción con el medio ambiente. En un sentido realmente no importa qué es qué, todo sistema inteligente debe estar situado de algún modo u otro para aspirar a ser una entidad inteligente.

La surgencia (emergence)

Al discutir donde reside la inteligencia en un programa de Inteligencia Artificial, Minsky puntualiza: *nunca hay un "corazón" en un programa y encontramos ciclos sin sentido y secuencias de operaciones triviales*. Es difícil señalar un sólo componente como el asiento de la inteligencia. No hay un homúnculo. En su lugar, la inteligencia surge de la interacción de los componentes del sistema. Sin embargo, la forma en la que *surge*, es muy diferente en los sistemas de la Inteligencia Artificial tradicional de la de los basados en el comportamiento.

En la Inteligencia Artificial tradicional, los módulos son de procesamiento o de adquisición de información, generalmente son de percepción, de planeación, de modelación del mundo, de aprendizaje, etc.

En la Inteligencia Artificial basada en el comportamiento, los módulos que se definen son productores de conductas. Los módulos podrían ser una conducta para evitar obstáculos, una conducta de detenerse, una conducta de control de ciclos, etc.

Aunque este dualismo entre los sistemas tradicionales y los basados-en-el-

comportamiento, luce bien, no es totalmente preciso. Los sistemas tradicionales difícilmente han sido conectados al mundo, y así la surgencia de una conducta inteligente es algo más que una expectativa en la mayoría de los casos, más que un fenómeno establecido. Recíprocamente, debido a las muchas conductas presentes en un sistema basado-en-el-comportamiento, y sus interacciones dinámicas individuales con el mundo, es difícil decir qué serie de acciones en particular fue producida por una conducta en particular. Algunas veces muchas conductas están operando simultáneamente, o están conmutando rápidamente [Horswill-Brooks88].

Durante muchos años se ha investigado sobre *surgencia* cuando se toca el tema de auto-organización. Dentro de los robots basados-en-comportamiento se está iniciando la investigación para caracterizar mejor la funcionalidad surgente. [Steels91] propone que el significado que una función se logra *indirectamente por la interacción de entre componentes más primitivos y con el mundo*.

Es difícil identificar el asunto de la inteligencia dentro de cualquier sistema, especialmente cuando la inteligencia se produce por las interacciones de muchos componentes. La inteligencia sólo puede ser determinada por la conducta total del sistema y por su relación con el medio ambiente.

Todo lo escrito hasta aquí sobre los sistemas basados en el conocimiento se puede resumir en las propias palabras de Brooks:

El mundo es su mejor modelo

El mundo aterriza la regresión filosófica.

La inteligencia está determinada por la dinámica de la interacción con el mundo.

La inteligencia está en el ojo del observador

La *funcionalidad inteligente* del sistema, tal como, percepción, planeación, modelación, aprendizaje, etc., *surge* de la interacción de los módulos de conductas simples o complejas.

En este capítulo se revisan los principios y fundamentos de aquella robótica que frecuentemente sigue el enfoque de la Inteligencia Artificial tradicional o de la distribuida y que aún cuando utilice modelos biológicos para instrumentar una conducta, es esencialmente una robótica no-reactiva



La cibernética como antecedente

En los años cuarenta y cincuenta, se inició una disciplina que tenía las mismas metas que la Inteligencia Artificial -la construcción de sistemas inteligentes útiles y el entendimiento de la inteligencia humana. Esta disciplina conocida como Cibernética se basaba en la teoría de control {Wiener61} y tenía características distintas de la Inteligencia Artificial tradicional.

La cibernética co-evolucionó con la teoría de control y la teoría estadística de la información. Es, en realidad, el estudio de las matemáticas de las máquinas, no en los términos de componentes funcionales de una máquina y de cómo están conectados y no en los términos de qué puede hacer, aquí y ahora una máquina individualmente, sino en los términos de todas las posibles conductas que una máquina individual puede producir. El énfasis es la caracterización de una máquina en términos de sus entradas y salidas, de hecho tratarla como una caja negra, aún cuando su trabajo interno fuera observable. Las herramientas de análisis son las ecuaciones diferenciales e integrales, que limitan a situaciones donde las condiciones de frontera no fueran rápidamente cambiantes (o discontinuas). Esto contrasta mucho con un sistema SITUADO que contiene con un mundo que cambia dinámicamente.

La cibernética surgió como el estudio del "control y comunicación en el animal y en la máquina". En ese entonces el modelo de computación era analógico. Típicamente se pensaba que las entradas y las salidas de la máquina al ser analizadas eran funciones continuas con derivadas razonables, y los mecanismos para el análisis y modelación automatizados eran como componentes analógicos. No había una noción de búsqueda simbólica -cualquier búsqueda era expresada en términos de optimización de una función. No había una noción de representación como una entidad abstracta manipulable tal como se encuentra en los enfoques de la Inteligencia Artificial tradicional.

Sorprendentemente la mayor parte de la investigación en Cibernética estaba dirigido a entender a los animales y a la inteligencia. Los animales eran modelados como máquinas, y de esos modelos, se esperaba coleccionar

5 La robótica no-reactiva

información de cómo los animales cambiaban sus conductas a través del aprendizaje, y de como eso conducía a una adaptación mejor del ambiente de todo el organismo. Prontamente se reconoció que un mecanismo y su ambiente deben ser modelados juntos para comprender la conducta producida por el organismo: una expresión de situacionalidad. Las herramientas de análisis de realimentación que eran usadas se concentraban en aspectos tales como estabilidad del sistema frente a la perturbación ambiental, y en particular en una homeostasis del sistema o habilidad para mantener ciertos parámetros dentro de ciertos rangos prescritos, sin importar las variaciones incontroladas del ambiente.

Con respecto a la INCORPORACION algunos experimentos estaban relacionados con robots físicos, [Walter50] describe robots construidos sobre principios cibernéticos que mostraban conductas de búsqueda de metas, homeostasis, y habilidades de aprendizaje. Sin embargo, la mayor parte de los modelos cibernéticos de organismos eran demostraciones abstractas de homeostasis.

La complejidad y habilidades de las máquinas incorporadas físicamente por Walter, caen en las puramente imaginarias de [Braitenberg84], tres décadas después.

Los factores limitantes en estos experimentos eran dobles; (1) la tecnología de construcción de pequeños robots auto-contenidos cuando los elementos computacionales eran miniatura (en términos relativos) de bulbos, y (2) la falta de mecanismos para describir abstractamente conductas a un nivel menor que la conducta completa, de tal manera que la implementación pudiera reflejar a esos componentes más simples. En la primera instancia los modelos de pensamiento estaban limitados por barreras tecnológicas, y en la segunda, la falta de ciertos componentes críticos de un modelo (organización en submódulos) restringía la habilidad de construir mejores implementaciones tecnológicas.

A mediados de los sesenta quedaba claro que el estudio de la inteligencia, estudio que surgía de los principios de la cibernética, si iba a ser exitoso,

necesitaba estar fundamentado más en sus niveles de abstracción y en sus herramientas de análisis que en cualquier otra cosa. Un buen ejemplo es el trabajo de Arbib.

La influencia de Wiener y Rosenblueth quienes abrigaban la esperanza de que los métodos cibernéticos pudieran dar un entendimiento de "todos los principios de coordinación e integración" que interrelacionan los subsistemas componentes del sistema nervioso humano, fue determinante para la falta de desarrollo de la Cibernética.

Los años que siguieron a la conferencia de Darmouth, en la que nació la Inteligencia Artificial, moldeó el campo de la misma de una manera que no ha cambiado significativamente. Los siguientes años, amplificaron el uso de la de abstracción alejándose con ello de la idea de SITUACIONALIDAD o conectividad con el mundo.

Al final de los años sesenta y principios de los setenta el proyecto del robot Shakey de [Nilsson84], en SRI, parecía consolidar las premisas de la Inteligencia Artificial clásica. Shakey era un robot móvil que habitaba un conjunto de cuartos especialmente preparados para él. Navegaba de cuarto en cuarto, tratando de satisfacer una meta, dada desde un teletipo. Dependiendo de la meta y de las circunstancias, navegaba alrededor de obstáculos que consistían en grandes bloques pintados y de marcas que se interponían en su camino o lo orientaban hacia un lugar deseado. Shakey tenía montada una cámara de televisión, en blanco y negro, como su sensor primario. Una computadora externa analizaba las imágenes y mezclaba las descripciones de lo que el robot veía con un modelo del mundo basado en cálculo de predicados de primer orden. Un programa de planeación (STRIPS), operaba esas descripciones simbólicas del mundo para generar una secuencia de acciones. Estos planes eran transformados a través de una serie de refinamientos, en invocaciones de acciones atómicas dentro de ciclos de realimentación ligados a operaciones de percepción también atómicas. Shakey utilizaba además de la cámara otros sensores tales odómetros (medidores de distancia a objetos).

5 La robótica no-reactiva

En ese entonces, Shakey fue considerado como un gran éxito, pues mostraba un sistema integrado que tenía movilidad, percepción, representación, planeación, ejecución y recuperación de errores.

Así el éxito de Shakey reafirmó la idea de confiar completamente en los modelos internos de una realidad objetiva externa. Esta es la metodología que se siguió y que pareció exitosa. Sin embargo, solo funcionaba debido a un ambiente muy cuidadosamente *ingenierado*. Veinte años más tarde, ningún robot móvil ha demostrado alcanzar todos los aspectos del desempeño de Shakey en un ambiente más general, tal como el de una oficina.

El hecho de que las habitaciones en las que operaba Shakey estuvieran desnudas, excepto por los grandes bloques de colores y las marcas, hacía que la representación de objetos fuera muy simple, lo que producía una visión robusta. La realidad objetiva del ambiente era muy simple, y el mapeo a un modelo interno de ella, también era muy plausible.

Por ese entonces en el MIT se montó una demostración de un robot que veía una escena que consistía de bloques apilados, y que luego construía una versión de ella usando un brazo-robot, véase [Winston72]. El programa para hacer esto, conocido como copy-demo, era específico del mundo de los bloques y reforzó la idea de que una descripción tridimensional completa del mundo podía ser extraída de una imagen visual a pesar del fracaso para reconocer figuras con textura rugosa o curvas o ambientes mal iluminados.

Ningún sistema de visión puede producir modelos del mundo fidedignos para algo cercano a la complejidad de las escenas reales del mundo --en nuestros días el reconocimiento de objetos es un área de investigación difícil y activa-- porque: (1) la visión por computadora nunca sepondrá al día y proporcionará tales modelos del mundo debido a la evidencia biológica, y (2) los modelos completos y objetivos de la realidad son irreales-- y por lo tanto los métodos de la Inteligencia Artificial que confían en tales modelos son irrealizables.

Un problema con los micromundos, dice Brooks, es que son aburridos. Los mundos de bloques eran los más populares de los micromundos y hay muy

5 La robótica no-reactiva

poco que pueda hacerse con ellos además de apilar bloques. Después de una confusión del trabajo inicial donde los *problemas* y *rompecabezas* particularmente difíciles fueron descubiertos y luego resueltos, de acuerdo con Sussman se volvió más y más difícil hacer algo nuevo dentro de ese dominio.

Había tres clases de opciones de repuesta intelectual a este espacio de problemas empobrecido:

- 1) Moverse a otro dominio con semántica igualmente simple, pero con nombres más interesantes que "bloque en", etc. Típicamente, no era esta la intención de los investigadores al hacerlo, pero de hecho muchos de ellos caían en esta trampa.
- 2) Construir una semántica más compleja en el mundo de los bloques y trabajar en los nuevos problemas que surgían. El problema con este enfoque es que las soluciones a los "rompecabezas" se volvieron tan específicas del dominio, que era difícil ver como se podrían generalizar a otros dominios.
- 3) Moverse a otro mundo más amplio. En particular, representar el conocimiento sobre el mundo diario, y luego construir solucionadores de problemas, sistemas de aprendizaje, etc., que operaran en mundos semánticamente más ricos.

Del último de estos enfoques ha fructificado posiblemente el subcampo más definido de la Inteligencia Artificial, conocido como *representación del conocimiento*. Tiene sus propios congresos. Tiene campos teóricos y prácticos. Pero está totalmente *desincorporado*. Concentra muchas de sus energías en las anomalías de los sistemas formales que nunca son usados para alguna tarea práctica.

[Brachman-Levesque85] tienen una colección de artículos del área. Los sistemas de representación del conocimiento descritos reciben sus entradas ya sea en forma simbólica o como salida de sistemas de lenguaje natural. Parece

5 La robótica no-reactiva

que la meta de los artículos es representar el *conocimiento* sobre el mundo. Sin embargo, repetimos está totalmente *desincorporado*. Hay muy pocos intentos de usar el conocimiento, salvo en las áreas de la física ingenua de Hayes, o de la física cualitativa de Kleer y Brown --pero nótese que estas áreas también son totalmente desincorporadas. Hay una suposición implícita de que algún día las entradas y salidas estarán conectadas con algo que las usará.

Desde el punto de vista pragmático, a pesar de las especificaciones mencionadas en los párrafos anteriores (ambientes estáticos y prefabricados, uso de las computadoras más poderosas disponibles) los robots operaban con una lentitud extrema, la mayor parte del tiempo de proceso era consumido en los fines perceptuales del sistema y en la construcción de modelos del mundo. El tiempo del proceso de cómputo dedicado a la planeación y la acción era menor. Como resultado de esto se creó un marco de trabajo para diseño y construcción de robots autónomos móviles, en el cual otros investigadores pudieran operar sin que tener que probar sus ideas en robots reales y aún sin tener acceso a los datos del robot. Este marco se ha denominado SMPA o *sensar-modelar-planear-actuar*, véase la figura 5-1.

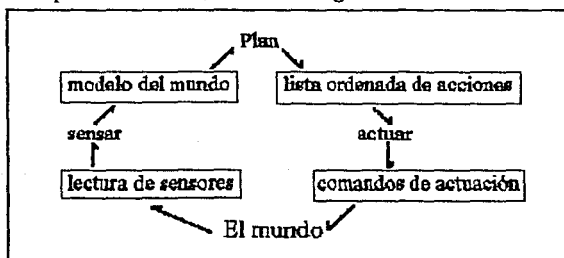


Figura 5.1 Modelo clásico de la Inteligencia Artificial para un control inteligente

Este marco influyó en la manera en la cual los robots eran construidos y los robots resultantes a su vez impusieron restricciones para la construcción de sus programas de control inteligente. Había por lo menos una suposición implícita en este trabajo temprano de robots móviles: una vez que se

5 La robótica no-reactiva

construyeran los casos simples, que operaran en ambientes estáticos, entonces las técnicas nuevas de cómputo resolverían los casos mas difíciles en medios dinámicamente activos. Ninguno de los primeros robots basados en el modelo SMPA tuvo éxito en ese sentido.

El tema crítico de la robótica clásica para nuestra búsqueda de principios planteada en la introducción recae en la planeación. En Inteligencia Artificial, y en particular en robótica, planeación se refiere a un conjunto de técnicas utilizadas para automatizar el proceso de seleccionar y llevar a cabo acciones para llevar un *ambiente* a un estado deseado.

La planeación clásica

En la *planeación clásica* se *codifica* un dominio de aplicación en términos de un conjunto de acciones primitivas, sus precondiciones y sus efectos caracterizados como predicados de estados. Un problema de planeación consiste de la descripción del dominio, más un estado inicial y un estado final (o meta). La planeación en este marco de trabajo consiste de buscar, a través de un espacio de ordenamientos de acciones, para con ello encontrar una secuencia de acciones que lleve del estado inicial al estado meta. A este conjunto de acciones ordenadas se le llama el plan.

El diseño de tales subsistemas ha dado lugar a la búsqueda de soluciones de muchos problemas en el campo de la Inteligencia Artificial, tales como representación y razonamiento sobre el tiempo; sobre la causalidad; sobre las intenciones; sobre las restricciones físicas; sobre la incertidumbre en la ejecución de la planeación; sobre el sentido y la percepción del mundo real; sobre el mantenimiento de creencias acerca de él; etc.

La planeación ha brillado poco en los dominios de aplicación en el mundo real por dos razones principales:

- a) el mundo cambia mientras la planeación está en progreso, y por lo tanto los planes parciales pueden resultar inútiles,
- b) los efectos inciertos de las acciones, al corregir el plan pueden

5 La robótica no-reactiva

ocasionar que éste falle para lograr su meta.

[Chapman87] ha mostrado que la planeación como programa presenta varias desventajas:

1. Computacionalmente existen problemas intratables;
2. El enfoque es inadecuado para caracterizar un mundo con eventos impredecibles
3. Requiere de los planes sean demasiado detallados
4. Fallan cuando se tiene que relacionar el texto del plan con situaciones concretas.

El problema de la planeación se dice que es, en el caso general, *indecidible*, y que aún en sus formas más simples puede ser computacionalmente *intratables*.

Desde el punto de vista clásico, la planeación se puede ver como un proceso de búsqueda, es decir, el planeador se organiza para primero definir un espacio de búsqueda y luego buscar un punto en él que se define como la solución. Los puntos del espacio son los estados de las aplicaciones en varios instantes de los operadores, eso implica que una solución se puede ver como la secuencia de operadores que recorren el espacio de búsqueda desde el estado inicial hasta el estado que satisface la meta. Así es posible tener planes parcialmente elaborados. Este camino, según comenta [Lara88], no es el que siguen los batracios. Esta razón entre otras excluye este tipo de robots no reactivos de nuestra consideración para nuestro objetivo.

La robótica reactiva es el tipo de arquitectura que hemos considerado adoptar para nuestra experimentación.

Su naturaleza no-representacional y su posición emergentista no sólo de conductas sino de control la hacen ideal para nuestro propósito



A mediados de los años 80, un gran número de investigadores comenzaron a preocuparse acerca del problema general de organización de los componentes para generar inteligencia reactiva. ¿Cuáles son los requerimientos para una inteligencia reactiva que responda a aspectos dinámicos del medio en los que robots móviles operen en escalas de tiempo similares a la de los humanos y a la de los animales. ¿Debería ser posible generar una conducta robusta aún contando con sensores inciertos y un medio impredecible, frente a un mundo cambiante. Algunas de las respuestas sobre esta organización:

1. La mayor parte de las cosas que las personas hacen día con día, no es solución de problemas o planeación sino más bien es una actividad rutinaria relativamente benigna pero ciertamente dinámica. Más aún las representaciones que un agente usa de objetos del Mundo, no necesariamente dependen de una correspondencia semántica como un símbolo que los agentes poseen, sino más bien puede ser definida por interacciones del agente con el Mundo. Agentes basados en estas ideas han logrado niveles de desempeño interesante y solamente fueron contruidos de circuitos combinatorios más una cierta cronometría.

2. Un observador puede legítimamente hablar de la creencia de los agentes y de los objetivos, aún cuando los agentes no necesariamente manipulen símbolos o estructuras de datos durante la etapa de ejecución. Un "aterizaje" formal semántico para el diseño de agentes puede descartarse.

3. Con objeto de probar realmente ideas de inteligencia, es importante construir agentes completos que operen en medios dinámicos usando sensores reales (proyecto de Dennett). Los modelos del mundo interno que tengan representaciones completas del mundo externo, además de ser imposibles de obtener, no son del todo necesarios para agentes que actúan de una manera competente. Muchas de las acciones de los agentes son completamente separables. La inteligencia coherente puede surgir de subcomponentes que interaccionando con el mundo.

Los nombres de la robótica reactiva.

No hay un término generalmente aceptado para describir este estilo de trabajo, alguna vez ha sido caracterizado como *planeación reactiva* para "deficientes mentales", y algunos han usado los nombres de seres robóticos como [BrooksR-FlynnA89] y criaturas artificiales como [Brooks91]. Los trabajos sobre visión activa o animada están relacionados con este tipo de investigación. Algunos investigadores se refieren a estos seres o criaturas como agentes, desafortunadamente el término también está usado por otros para referir algunos componentes independientes de inteligencia dentro de una criatura física, véase el caso de Minsky. Algunas veces la manera de nombrarlos es comportamiento basal. Brooks propone, simplemente llamar a las entidades en discusión, robots o robots basados-en-el-comportamiento.

Recientemente ha habido una tendencia a tratar de integrar el razonamiento simbólico de tipo tradicional *sobre* un sistema puramente reactivo tanto en robots como en simulación. La idea es que los sistemas reactivos manejan asuntos de tiempo real o esten embebidos en el mundo mientras que los sistemas *deliberativos* hacen el duro trabajo tradicional, imaginados por los sistemas de Inteligencia Artificial. Brooks piensa que estos puntos de vista sufren del bien conocido efecto de horizonte: han aprovechado la ligera mejoría en el desempeño global del sistema usando un componente reactivo. Lo que han hecho es simplemente empujar en el desempeño los límites del sistema razonamiento un poco más allá, en el futuro.

[Brooks86] notó que las *arquitecturas inteligentes de control* de robots estándares, vistas como colecciones *entubadas* de módulos funcionales, causaban problemas de robustez, constructibilidad y verificabilidad. El autor mencionado sugirió una arquitectura novedosa, llamada arquitectura de inclusión (subsumption), que enfatiza la construcción del control inteligente en *capas* de conductas que-logran-tareas muy en el espíritu de [Braitenberg84]. Brooks, con estas ideas, ha construido un buen número de sistemas robóticos que exhiben conductas autónomas y que parecen actuar de una manera robusta, inteligente en ambientes complejos; y sólo cuentan internamente con reglas de control oportunísticas simples.

No Existe

Página

Los principios de los sistemas basados en el comportamiento

Todas las investigaciones se desarrollan dentro de las restricciones impuestas por ciertos principios. Algunas veces están explícitas y otras veces están implícitas. En los párrafos siguientes se delinearán explícitamente los principios de los sistemas basados en el comportamiento.

El primer conjunto de principios define el *dominio del trabajo*.

- La meta es estudiar agentes inteligentes autónomos completamente integrados.
- Los agentes deberán estar incorporados como robots móviles, situados en mundos sin modificar, que se encuentran alrededor del laboratorio. Esto confronta el aspecto de la incorporación. Los ambientes elegidos son por conveniencia, aunque Brooks y sus seguidores resisten la tentación de cambiar los ambientes de cualquier forma para los robots.
- Los robots debieran estar operando igualmente bien cuando visitantes o trabajadores caminen a través de su área de trabajo, cuando se reorganice la decoración, cuando la iluminación u otras condiciones ambientales cambien, y cuando sus sensores y actuadores deriven sus calibraciones. Esto confronta el aspecto de situado.
- Los robots debieran operar a escalas de tiempo medidas en las escalas de tiempo usadas por las personas. Esto también confronta el aspecto de Situacionalidad.

El modo específico de computación usado, no estaba basado originalmente en modelos biológicos. Se fue arribando a él por intentos de refinamientos continuos para programar un robot que evitara colisiones, reactivamente, en un ambiente poblado de personas, [Brooks 86]. Sin embargo, para establecer los principios utilizados en el modelo de computación, es claro

que comparte ciertas propiedades con modelos de como los sistemas neurológicos se arreglan (organizan, ordenan). Es importante enfatizar que sólo comparten ciertas propiedades. El modelo de computación de Brooks no está intentado como un modelo realista de como trabajan los sistemas neurológicos. Al modelo de computación le han llamado la arquitectura *subsumption* y su propósito es programar agentes inteligentes, situados e incorporados.

Los principios de *computación* son:

- La computación está organizada como una red asíncrona de elementos computacionales activos [Brooks86], en una red de topología fija de conexiones unidireccionales.
- Los mensajes son enviados sobre las conexiones, no tienen una semántica implícita --son pequeños números (típicamente 8 ó 16 bits, pero en algunos robots sólo 1 bit) y sus significados son dependientes de la dinámica diseñada en el emisor y en el receptor.
- Los sensores y los actuadores están conectados a esta red, generalmente a través de "buffers" asíncronos de dos lados.

Estos principios conducen a ciertas consecuencias. En particular:

- El sistema puede tener estados --no está todo restringido a ser reactivo.
- Puede haber apuntadores y estructuras de datos manipulables aunque son muy difíciles de implementar.
- Cualquier espacio de búsqueda debe estar bien acertado en tamaño, los nodos de búsqueda no pueden ser creados y destruidos dinámicamente durante el proceso de búsqueda.
- No hay una separación implícita de datos y computación, ambas están distribuidas sobre la misma red de elementos.

Considerando las observaciones biológicas presentadas en el capítulo 2, parece que ciertas propiedades se incorporan en la forma en que los robots se

programan dentro de un modelo de computación. En la experimentación con robots basados en el comportamiento se han observado los siguientes principios de *organización*:

- No se mantiene un modelo central del mundo. Todos los datos están distribuidos en muchos elementos computacionales.
- No hay un sitio central de control.
- No hay una separación entre el sistema perceptual, el sistema central y el sistema de actuación. Secciones de la red puede realizar más de una de estas funciones.
- La competencia conductual del sistema se mejora agregando redes de conductas específicas a la red existente. Se llama a este proceso adición de capas (*layering*), que es una analogía cruda y simplista del proceso evolutivo. Como con la evolución, en cada etapa del desarrollo los sistemas se prueban a diferencia de la evolución, hay un proceso de depuración. Cada una de las capas es una sección de la red que produce una conducta en sí misma.
- No hay un arreglo jerárquico --no hay una noción de un proceso invocando a otro a la manera de subrutina. Más bien, las redes están diseñadas para que los cálculos requeridos simplemente estén disponibles en la línea de entrada cuando se necesiten. No hay sincronización explícita entre un productor y un "consumidor" de mensajes. Los "buffers" de recepción de mensajes pueden ser reemplazados por nuevos mensajes antes de que un "consumidor" vea el mensaje anterior. No es atípico que un productor de mensajes envíe 10 mensajes por cada uno es examinado por el receptor.
- Las capas, o conductas, todas corren en paralelo. Puede requerirse un mecanismo de resolución de conflictos cuando conductas diferentes traten de comandar a los mismos actuadores.

incluyen memoria EEPROM y RAM, es posible usar un gran número de sensores que requieren servicios de interrupción, calibración local y envío de mensajes. Los microprocesadores pueden reducir significativamente la complejidad del alambreado y así dar servicio, *in situ*, a un grupo local de sensores (p.e. todos ellos en una sola pierna de un robot).

Las características de robots físicos construidos bajo este nuevo enfoque se discuten a continuación

La reactividad

La primera demostración de una arquitectura de inclusión se hizo con el robot *Allen* [Brooks86]. Este era casi totalmente reactivo, usaba lecturas de sonar para mantenerse alejado de la gente y de otros obstáculos móviles y sin colidir con obstáculos fijos. El robot se movía hacia el lugar elegido mientras la capa reactiva de nivel inferior cuidaba de evitar los obstáculos. Este primer robot combinaba capacidades no reactivas con reactivas. Pero el punto importante es que usaba exactamente las mismas clases de mecanismo computacional para hacer ambas. Al observar la red de capas combinadas se nota que no hay una partición obvia entre los componentes de nivel inferior y superior. Sin embargo, había una diferencia funcional entre las dos capas.

El segundo robot *Herbert* [Connell89], con enfoque reactivo. Hacía uso de un rastreador láser que le permitía encontrar objetos similares a una lata de refresco. Tenía sensores infrarrojos de proximidad que le permitían navegar siguiendo paredes y pasar a través de dinteles de puertas. También tenía una brújula magnética para mantener un sentido global de orientación y un grupo de sensores en un brazo, que era lo suficientemente confiable para sostener latas de refrescos. La tarea dada a *Herbert* era la de explorar en busca de, latas de refresco, encontrarlas y llevarlas al lugar del que había partido. *Herbert* mostró ser un robot confiable para este objetivo.

Se decidió programar *Herbert* de tal manera que no debía permanecer en un estado más de tres segundos, y que no debía haber comunicación interna

entre los módulos generadores de conductas. Cada uno de los módulos estaba conectado a sensores en el lado de las entradas, y a una red de arbitraje de prioridad fija en el lado de la salida. La red de arbitraje dirigía a los actuadores.

En muchos casos para llevar a cabo sus tareas *Herbert* hacía uso del propio mundo como su mejor modelo y como un medio de comunicación. Por ejemplo, el buscador de latas de refresco dirigía al robot de tal manera que su brazo quedara alineado frente a la lata de refresco. Pero no comunicaba al controlador del brazo que había un objeto listo para ser agarrado. En ese caso las conductas del brazo monitoreaban a los cifradores de las ruedas, y cuando este monitoreo notaba que no había movimiento del cuerpo, se iniciaban los movimientos del brazo. Esto a su vez disparaba otras conductas, las que permitían al robot alcanzar la lata.

La ventaja de este enfoque es el no contar con una predicción de lo que pasaría a continuación lo que se traduciría en un sistema de control:

- 1) oportunista, si circunstancias fortuitas se presentaban y
- 2) rápido para responder a circunstancias cambiantes, como el caso de un objeto que se acerca en una dirección de colisión.

Un ejemplo de como las conductas del brazo se producen en cascada una tras otra, es el acto de agarrar la lata de refresco. La garra tiene un reflejo de asir que opera siempre que algo interrumpe el rayo infrarrojo entre los dedos. Cuando el brazo localiza una lata con sus sensores locales, simplemente dirige la mano para que dos dedos queden alineados a los lados de la lata. Solo entonces la mano agarra la lata. Dado este arreglo, es posible que una persona de al robot una lata. Tan pronto como era asida, el brazo se retraía --no importa en este caso si la lata había sido dada intencionalmente o había aparecido mágicamente. Este mismo oportunismo entre conductas permite que el brazo se adapte automáticamente a una variedad de muebles sobre los que puede encontrar exitosamente la lata.

entre los módulos generadores de conductas. Cada uno de los módulos estaba conectado a sensores en el lado de las entradas, y a una red de arbitraje de prioridad fija en el lado de la salida. La red de arbitraje dirigía a los actuadores.

En muchos casos para llevar a cabo sus tareas Herbert hacía uso del propio mundo como su mejor modelo y como un medio de comunicación. Por ejemplo, el buscador de latas de refresco dirigía al robot de tal manera que su brazo quedara alineado frente a la lata de refresco. Pero no comunicaba al controlador del brazo que había un objeto listo para ser agarrado. En ese caso las conductas del brazo monitoreaban a los cifradores de las ruedas, y cuando este monitoreo notaba que no había movimiento del cuerpo, se iniciaban los movimientos del brazo. Esto a su vez disparaba otras conductas, las que permitían al robot alcanzar la lata.

La ventaja de este enfoque es el no contar con una predicción de lo que pasaría a continuación lo que se traduciría en un sistema de control:

- 1) oportunista, si circunstancias fortuitas se presentaban y
- 2) rápido para responder a circunstancias cambiantes, como el caso de un objeto que se acerca en una dirección de colisión.

Un ejemplo de como las conductas del brazo se producen en cascada una tras otra, es el acto de agarrar la lata de refresco. La garra tiene un reflejo de asir que opera siempre que algo interrumpe el rayo infrarrojo entre los dedos. Cuando el brazo localiza una lata con sus sensores locales, simplemente dirige la mano para que dos dedos queden alineados a los lados de la lata. Solo entonces la mano agarra la lata. Dado este arreglo, es posible que una persona de al robot una lata. Tan pronto como era asida, el brazo se retraía --no importa en este caso si la lata había sido dada intencionalmente o había aparecido mágicamente. Este mismo oportunismo entre conductas permite que el brazo se adapte automáticamente a una variedad de muebles sobre los que puede encontrar exitosamente la lata.

Para regresar al lugar del que había partido antes de asir la lata, *Herbert* usaba un truco. Las rutinas de navegación podrían llevar reglas implementadas tales como: "Cuando pases a través de la puerta, da vuelta a la izquierda". Estas reglas estaban condicionadas a la preparación de los modelos de la mano. Cuando el robot estaba fuera de límites sin lata en su mano, ejecutaba eficazmente un conjunto de reglas. Diseñando cuidadosamente las reglas, *Herbert* estaba garantizado con una confiabilidad razonable, para retrazar su trayectoria.

El punto de traer a colación esta discusión sobre el robot *Herbert* es doble:

- Mostrar cómo una conducta compleja, aparentemente dirigida por la meta, intencional, ocurre en un sistema que no tiene estado interno de largo alcance y que carece de comunicación interna entre sus módulos.
- Mostrar que es muy fácil para un observador el atribuir a un sistema una estructura interna más compleja de la realmente tiene. *Herbert* parece estar haciendo cosas como planeación de trayectorias y construcción de mapas, aún cuando no lo hace.

La representación

El trabajo de Brooks, se critica frecuentemente, por el hecho de abocarse totalmente sin representación del mundo a un robot basado-en-conductas. Esta crítica es inválida aunque en realidad se opone al concepto de representación de la Inteligencia Artificial tradicional y rechaza las representaciones explícitas de metas dentro de la máquina, acepta que pueden haber representaciones que sean modelos parciales del mundo --de hecho hemos mencionado que *las capas individuales extraían únicamente aquellos aspectos del mundo que encuentra relevantes* [Brooks 91a]. Dentro de este contexto computacional la forma que estas representaciones toman, dependerá de la tarea para la que se utilizan las representaciones.

[Mataric90] introdujo el concepto de representaciones activo-constructivas

Para regresar al lugar del que había partido antes de asir la lata, Herbert usaba un truco. Las rutinas de navegación podrían llevar reglas implementadas tales como: "Cuando pases a través de la puerta, da vuelta a la izquierda". Estas reglas estaban condicionadas a la preparación de los modelos de la mano. Cuando el robot estaba fuera de límites sin lata en su mano, ejecutaba eficazmente un conjunto de reglas. Diseñando cuidadosamente las reglas, Herbert estaba garantizado con una confiabilidad razonable, para retrasar su trayectoria.

El punto de traer a colación esta discusión sobre el robot Herbert es doble:

- Mostrar cómo una conducta compleja, aparentemente dirigida por la meta, intencional, ocurre en un sistema que no tiene estado interno de largo alcance y que carece de comunicación interna entre sus módulos.
- Mostrar que es muy fácil para un observador el atribuir a un sistema una estructura interna más compleja de la realmente tiene. Herbert parece estar haciendo cosas como planeación de trayectorias y construcción de mapas, aún cuando no lo hace.

La representación

El trabajo de Brooks, se critica frecuentemente, por el hecho de abocarse totalmente sin representación del mundo a un robot basado-en-conductas. Esta crítica es inválida aunque en realidad se opone al concepto de representación de la Inteligencia Artificial tradicional y rechaza las representaciones explícitas de metas dentro de la máquina, acepta que pueden haber representaciones que sean modelos parciales del mundo --de hecho hemos mencionado que "las capas individuales extraían únicamente aquellos aspectos del mundo que encuentra relevantes"[Brooks 91a]. Dentro de este contexto computacional la forma que estas representaciones toman, dependerá de la tarea para la que se utilizan las representaciones.

[Mataric90] introdujo el concepto de representaciones activas-constructivas

para ser empotradas (embedded) en un robot basado en un sonar, Toto, que ambulaba por ambientes de oficina construyendo mapas basados en etiquetas y luego usaba ese mapa para ir de un lugar a otro. Sus representaciones estaban totalmente descentralizadas y no eran manipulables, no había un control central que construyera, mantuviera o usara los mapas. El mapa mismo era una estructura activa que hacía necesario el cómputo para cualquier planeación de trayectoria que el robot necesite hacer.

Las capas inferiores de control permitían que *Toto* deambulara siguiendo fronteras (tales como paredes). El mapa se representa como una gráfica. Los nodos de la gráfica son elementos computacionales que son a su vez redes pequeñas idénticas que corresponden a distintas máquinas de estados finitos aumentadas.

Los nodos libres de la gráfica se arbitran y distribuyen ellos mismos, de una manera totalmente local, para representar una nueva etiqueta, y establecer ligas topológicas con nodos físicamente vecinos. Estos nodos "recuerdan" al robot dónde se encuentra físicamente ubicado. Cuando una capa superior desea que el robot vaya a alguna marca conocida, simplemente "estimula" en una forma particular en el lugar del mapa al que desea ir. A medida que se mueve el robot también lo hace su representación.

El trabajo de Mataric enfatiza los siguientes aspectos de los sistemas basados en el comportamiento:

- Tales sistemas pueden hacer predicciones acerca de lo que pasará en el mundo y por lo tanto tendrán expectativas.
- Tales sistemas pueden hacer planes -pero no en la forma tradicional de la Inteligencia Artificial- véase [AgreP-ChapmanD87], para un análisis de este aspecto.
- Tales sistemas pueden tener metas -véase el capítulo 7 para otra forma de implementar metas dentro de este punto de vista.

-Todo lo anterior se puede hacer sin manipular representaciones simbólicas centrales o globales.

La complejidad de los sistemas

¿Pueden los sistemas ya desarrollados como robots escalar a sistemas arbitrarios complejos? Esta es una pregunta que no se puede contestar afirmativamente en el momento actual. Sin embargo, hay un número de dimensiones a través de las cuales la pregunta de escalamiento anterior todavía se puede hacer:

- Estas implementaciones pueden trabajar todavía bien a medida que el ambiente se vuelve más complejo
- Se pueden manejar números grandes de sensores y actuadores con este el enfoque.
- Siempre se puede encontrar alguna funcionalidad a medida que se agregan más y más capas o conductas.

El proceso de aprendizaje

La evolución nos enseña que hay un compromiso entre lo que sabemos a través de nuestros genes y lo que debemos encontrar por nosotros mismos a medida que nos desarrollamos. Podemos esperar a ver un compromiso similar para los robots basados-en-el-comportamiento.

En cuanto a aprendizaje se refiere hay al menos, cuatro clases de cosas que se pueden aprender:

1. Representaciones del mundo que pueden ayudar en alguna tarea.
2. Aspectos relativos a parámetros de sensores y actuadores (algunas veces se llama calibración).
3. Las formas en que las conductas individuales debieran interactuar.
4. La creación de nuevos módulos conductuales.

-Todo lo anterior se puede hacer sin manipular representaciones simbólicas centrales o globales.

La complejidad de los sistemas

¿Pueden los sistemas ya desarrollados como robots escalar a sistemas arbitrarios complejos? Esta es una pregunta que no se puede contestar afirmativamente en el momento actual. Sin embargo, hay un número de dimensiones a través de las cuales la pregunta de escalamiento anterior todavía se puede hacer:

- Estas implementaciones pueden trabajar todavía bien a medida que el ambiente se vuelve más complejo
- Se pueden manejar números grandes de sensores y actuadores con este enfoque.
- Siempre se puede encontrar alguna funcionalidad a medida que se agregan más y más capas o conductas.

El proceso de aprendizaje

La evolución nos enseña que hay un compromiso entre lo que sabemos a través de nuestros genes y lo que debemos encontrar por nosotros mismos a medida que nos desarrollamos. Podemos esperar a ver un compromiso similar para los robots basados-en-el-comportamiento.

En cuanto a aprendizaje se refiere hay al menos, cuatro clases de cosas que se pueden aprender:

1. Representaciones del mundo que pueden ayudar en alguna tarea.
2. Aspectos relativos a parámetros de sensores y actuadores (algunas veces se llama calibración).
3. Las formas en que las conductas individuales debieran interaccionar.
4. La creación de nuevos módulos conductuales.

7 Los enfoques representativos de la robótica reactiva

En este capítulo se presentan dos variantes de la arquitectura de robots reactivos.

Una que corresponde al enfoque jerárquico que ha sido adoptada para nuestro trabajo experimental. Y la otra que corresponde a la robótica reactiva heterárquica, la que a pesar de su atractivo emergentismo de control no es la arquitectura de robótica reactiva que nos conviene para el propósito de la tesis ya que viola el principio fundamental jerárquico que es nuestra referencia biológica. Sin embargo, la importancia de este último enfoque es grande, por la posibilidad de construir robots jerárquicos-heterárquicos



Introducción

[Brooks91] ha propuesto la construcción de robots inteligentes separando el razonamiento y el control, de tal manera que sin tener un módulo de razonamiento, como sucede con el enfoque tradicional, se pueda armar un control que produzca una conducta global considerada inteligente. En el enfoque tradicional de la inteligencia artificial se denomina planeación al proceso de razonamiento que utiliza la representación que el agente tiene de las acciones, metas y eventos, este proceso determina la secuencia de acciones que conducirá al logro de las metas [Georgeff89]. En el presente enfoque de la Inteligencia Artificial las arquitecturas son totalmente distintas y comparten las características siguientes:

- ✎ acoplamiento directo entre la percepción y a acción;
- ✎ control distribuido y descentralizado;
- ✎ interacción directa con el ambiente;
- ✎ mecanismos intrínsecos para contender con recursos limitados y conocimiento incompleto;
- ✎ funcionalidad surgente.

Se persigue la idea de funcionalidad surgente basada en una forma distribuida de control a través de las interacciones paralelas locales interconectadas y entre el medio ambiente y las conductas.

La organización del control se puede establecer de manera jerárquica o heterárquica. La dificultad de describir el tipo de control requerido ha propiciado la creación de diversas especificaciones. A continuación se describen dos modelos computacionales que corresponden a cada una de estas categorías.

Arquitecturas jerárquicas

Un trabajo fundamental para nosotros es el de [Kaelbling90], su versión de componente primitivo de un sistema reactivo es la de un mediador. La representación que esta autora da de esta unidad se puede ver en la figura 7.1. En ella se puede ver que el mediador está formado por dos subsistemas en serie: el perceptual y el de acción. Sin embargo, no se trata de un sistema abierto, la acción también realimentada al sistema perceptivo, lo mismo se manifieste que no. En el primer caso el proceso se puede identificar como un mecanismo de ATENCIÓN y en el segundo caso como un mecanismo de ENFOQUE.

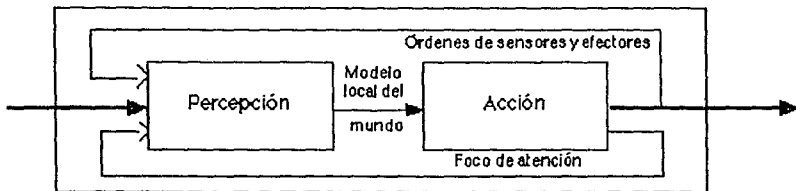


Figura 7.1 Componente primitivo (mediador) de la arquitectura jerárquica de Kaelbling

La arquitectura jerárquica propuesta por Kaelbling se basa en la idea de que un mediador puede estar formado por mediadores, de tal manera que el robot reactivo completo hasta su más alto nivel es un mediador, lo que implica una estructura jerárquica de mediadores como la de la figura 7-2.

El control del robot se divide jerárquicamente en términos de la composición de conductas. Donde cada conducta se define como un procedimiento que mapea un conjunto de entradas (en este caso las salidas del módulo perceptual) en un conjunto de salidas para los efectores del sistema. La composición de conductas las realiza el mediador mismo.

7 Los enfoques representativos de la robótica reactiva.

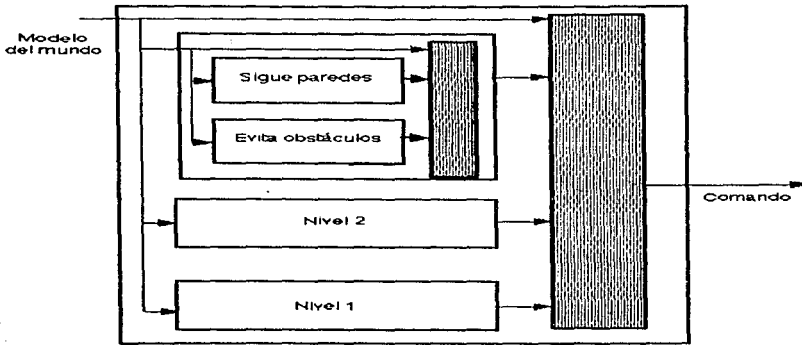


Figura 7.2 Diagrama jerárquico de mediadores, cada módulo genera una conducta siguiendo el esquema presentado en la figura 7.1 y los cuadros sombreados con los mediadores

También tomado de Kaelbling transcribimos a continuación una versión en pseudocódigo de la conducta de sigue-paredes, se supone que en ella cada mediador consiste de un comando motor y un comando sensor, y sólo puede activar o consultar un sensor a la vez:

```

comando-sensor := si info-izq-debil entonces activa(sensor-izq)
                  de otro modo si info-der-debil entonces activa(sensor-der)
                  de otro modo *nada*

comando-motor := si localiza-pared-izq(k) y localiza-pared-der(k)
                  entonces activa-servo(medio) de otro modo *no-comando*
    
```

Nosotros hemos adoptado la arquitectura jerárquica antes mencionada como el esquema fundamental para la simulación de nuestros robots. Esta decisión obedece al deseo de mantenernos aunque sea en su forma más general dentro de un esquema de organización biológica: recuérdese la organización jerárquica que tiene el sistema nervioso de los animales.

Arquitecturas heterárquicas

La estrategia biológica para elegir la conducta óptima en los animales simples está *alambrada*, por el proceso de selección natural, al igual que en las creaturas artificiales como las construidas por Brooks. Sin embargo, los seres vivos y las creaturas complejas necesitan de un mecanismo de selección de conducta más flexible que base su decisión en el estado motivacional interno así como en circunstancias externas. [Langton88] ha propuesto que la *funcionalidad surgente* es un principio importante en la organización de los seres vivos y que una estrategia más exitosa para que un sistema tenga la habilidad de reaccionar flexible y robustamente a un ambiente complejo y dinámico es diseñar una funcionalidad de una manera ascendente, haciendo que ella emerja como un efecto global colateral de las interacciones locales e intensivas entre los componentes del organismo.

[Maes91] establece que el proceso de selección se puede *modelar* mediante una aproximación ascendente incremental, utilizando una dinámica de activación/inhibición de las diversas conductas que se pueden elegir. No hay una asignación de pesos al estilo cognitivo, ni hay jerarquías burocráticas. De manera que el proceso de selección entre una variedad de conductas para decidir cuál de ellas ha de *ganar* (y así obtener el control de los actuadores o *músculos* de la creatura) se debe basar en mecanismos ascendentes y se debe modelar como una *propiedad surgente* de un proceso paralelo no-jerárquico ni preprogramado.

La especificación por sí sola de la conducta del agente, no explica la funcionalidad desplegada cuando el agente está operando. En su lugar, la funcionalidad, en un alto grado, está fundamentada en las propiedades del medio ambiente. De manera tal que además de considerar dinámicamente al ambiente, se explotan sus características, para que sirvan al funcionamiento del sistema. Esto implica que para lograr un objetivo hay que indicar (programar) al agente y además encontrar un ciclo de interacción que involucre al sistema y al ambiente que converja hacia la meta.

El mecanismo de selección de conductas se basa en la selección del *estado motivacional* interno de la creatura así como en las circunstancias externas.

7 Los enfoques representativos de la robótica reactiva.

De manera que cada creatura tiene asociado un conjunto de conductas que pueden tener las características siguientes:

- ✎ un nivel de activación;
- ✎ condiciones o estímulos que se tienen que observar para que una conducta se ejecute;
- ✎ un umbral de disparo, si el nivel de activación de la conducta sobrepasa este umbral, la conducta se coloca en la lista de las conductas activas o candidatas a ser ejecutadas;
- ✎ motivaciones;
- ✎ condiciones perceptuales, que corresponden a las lecturas de los sensores (o bien a la "fuerza" de la percepción)

Las conductas y sus motivaciones asociadas que Maes maneja en un experimento de simulación que ha hecho, se muestran en la tabla 7-1.

conductas	motivaciones
evitar-obstáculos	seguridad
explorar	curiosidad
pelear	agresión
huir	miedo
comer	hambre
dormir	lasitud, cansancio
beber	sed
acercarse a un objeto	
acercarse a la comida	
ir hacia el agua	

Tabla 7-1 Conductas y motivaciones simuladas por Maes en sus creaturas con control heterárquico.

Las conductas de una creatura están organizadas en una red de enlaces de distinta naturaleza: predecesores, sucesores y antagonismos, véase la figura 7-3. Con estas ligas se pretende representar aquellas subconductas que son innatas.

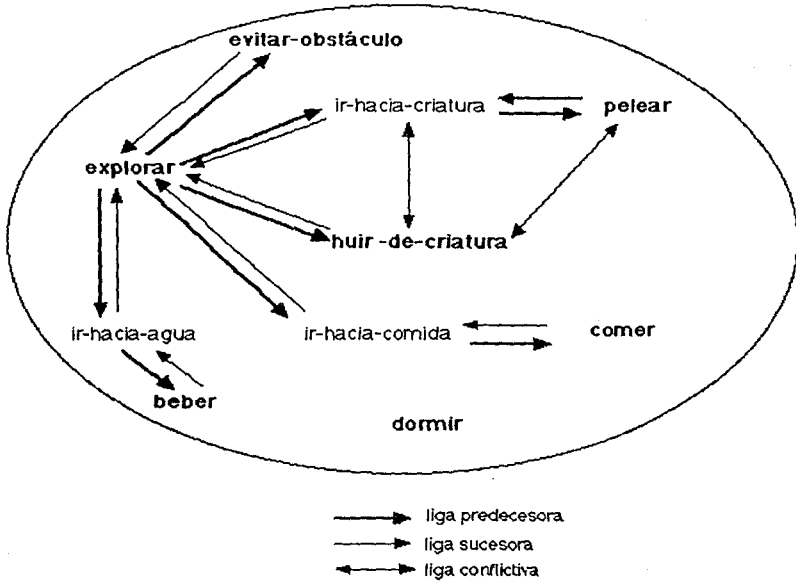


Figura 7.3 Diagrama de las relaciones entre las conductas

Las conductas están asociadas directamente con una o más motivaciones. Una vez que la conducta se ejecuta el nivel de las motivaciones se decrementa. La última de la cadena de conductas es consumatoria. Las conductas que no están asociadas directamente con algún consumo, se dice que son apetitivas. Estas conductas son las precondiciones de conductas consumatorias.

Puede haber dualidad en las conductas es decir, para alguna motivación la conducta puede ser apetitiva, y para otra puede ser consumatoria.

La naturaleza de la interacción de los robots reactivos simulados exige un medio físico en el que los *recursos energéticos* se agoten y que los robots

7 Los enfoques representativos de la robótica reactiva.

cuenten con un proceso informativo sobre la estimación de tales recursos y de la misma manera se cuente con los recursos energéticos internos agotables y detectables por el robot.

La idea intuitiva de la red y de las ligas entre conductas es que estas se utilicen para aumentar o disminuir la activación de sus vecinas inmediatas

Los resultados de este enfoque son bastantes alentadores puesto que las creaturas simuladas, además de compartir las características mencionadas al inicio del capítulo, muestran algunas de las características plásticas de los animales como son el *desplazamiento conductual*, la *fatiga* y la *habitación*. Sin embargo, el enfoque tiene la desventaja de que se tiene que *experimentar* sin una guía racional, cambiar constantemente los valores de los niveles de activación y los umbrales para obtener los resultados mencionados, no quedando al diseñador más remedio que hacer un ajuste intuitivo de los parámetros.

Existen numerosos síntomas de autonomía que pueden conducir a aceptar que un sistema dado es cognitivo (intencional).

Este capítulo enfoca ya la experimentación con robots cognitivos implícitos de tipo reactivo en su modalidad navegante, ya que ella ofrece el requisito de intencionalidad que se necesita para tal propósito.

Aquí se reportan el detalle de la arquitectura robótica utilizada y los experimentos conducidos con ella así mismo se presentan los principios cognitivos encontrados



Introducción

La navegación reactiva que hemos elegido para nuestra experimentación, es una forma de control que está caracterizada por una *filosofía* de tipo estímulo-respuesta a un *mundo*. Comportamientos complejos surgen como una combinación de respuestas simples, de bajo nivel, a los estímulos que el *mundo* aporta. Típicamente, esto implica la descomposición de tareas en una colección de subtareas distribuidas paralelamente.

El sentido propositivo de la navegación.

La navegación sólo puede considerarse como tal si existe una intención (*navegar a* o *navegar hacia*). En nuestro caso avanzar en una dirección o alcanzar o huir de un objeto-meta. Como se verá más tarde, este componente es una condición necesaria para asignar cognitividad a nuestros robots.

Autonomía

El enfoque enactivo, [Varela88], de la Inteligencia Artificial arguye que la autonomía y no la tenacidad el logro de la en la solución de problemas es la propiedad clave que define la noción intuitiva de *creatura inteligente*. Para construir una entidad inteligente que actúe autónomamente, primero debemos entender los atributos de un sistema que nos hacen llamarlo autónomo, (al menos en la práctica). La presencia de estos atributos proporciona a los sistemas autónomos la apariencia de *indeterminista*, aunque sabemos que ésta puede estar presente tanto en los sistemas vivientes como en artefactos determinísticos. También se argumenta que la autonomía significa tener las clases correctas de metas y la habilidad para elegir las metas de un conjunto existente de ellas, la autonomía no necesariamente está en el crear nuevas metas.

Según [Covirgaru-Lindsay90] se puede tener un sistema determinístico al cual le podemos asignar autonomía en función de las características que se asocien a las metas

En la Tabla 8.1 se resumen algunas de las características contrastantes de las metas que pueden ser relevantes para identificar a un agente como autónomo.

Pares de conjuntos contrastantes de metas		relevante	autónomo
Interconstruida	Adquirida	no	----
Endógena	Exógena	no	----
Simple	Múltiple	sí	Múltiple
Submetas	Metas de alto nivel	sí	Alto-nivel
Alcanzable	Homeostática	sí	Homeostática

Tabla 8.1. Esta tabla muestra las dimensiones a lo largo de las cuales se definen los tipos de metas y su relevancia en la definición de autonomía

La autonomía es una propiedad subjetiva; esto es, es una propiedad de una descripción del sistema por una “tercera persona”: no es una propiedad objetiva que existe independientemente de los observadores, es una propiedad relativa al sujeto que la observa.

Un sistema dirigido por metas será percibido como autónomo en tanto que:

- ☛ seleccione tareas (metas de alto nivel) que lo dirijan en cualquier instante;
- ☛ el sistema exista como tal durante un período de tiempo que sea relativamente largo comparado con el tiempo que requiere para lograr una meta;
- ☛ sea robusto, permanece viable en un ambiente variable;
- ☛ alguna de sus metas sean homeostáticas;
- ☛ haya metas que estén activas (evocadas pero no logradas);
- ☛ interaccione con su medio ambiente en un modo de procesamiento de información;
- ☛ exhiba alguna variedad de repuestas complejas, incluyendo movimientos adaptativos;
- ☛ su atención a los estímulos sea selectiva;
- ☛ ninguna de sus funciones, acciones o decisiones necesiten estar totalmente controladas por un agente externo; y
- ☛ una vez que el sistema empiece a funcionar, no necesite más programación.

Como se caracterizó, la autonomía no entraña un nivel de complejidad extremo o inteligencia. Los organismos simples vivos son autónomos aún cuando su inteligencia (medida por la complejidad de su repertorio computacional) esté limitada. Una máquina que juega ajedrez exhibe la solución de problemas (limitados a una tarea) pero no es autónoma según el criterio anterior. También diríamos que no es inteligente según el saber popular debido a que la inteligencia, como generalmente se concibe, requiere autonomía.

La cognitividad de los robots jerárquicos navegantes.

Una vez alcanzado este punto y la exploración de la cognitividad maquinística nos queda claro que la cognitividad de los robots navegantes no solamente debe postularse como una cuestión de principios propositivos (de navegación, por ejemplo).

Postulamos que hay principios cognitivos implícitos que pueden asignarse en tales robots, a principios de organización de los componentes internos que resuelven un problema de supervivencia en un ambiente dado.

La supervivencia puede medirse como una "figura de mérito" de la adaptación que hace un programador con los componentes de un robot o de autoadaptación del robot. Esta puede tomar distintas formas, una de ellas como el mantenimiento de la integridad física y otra como la permanencia en la búsqueda exitosa de propósitos.

Para medir la primera, en nuestros experimentos, se utilizó la forma del inverso de la frecuencia de la intervención de la actividad táctil de un robot, y para la segunda, el tiempo de permanencia exitosa del robot en el ambiente, en la consecución de sus objetivos navegacionales (viajar en cierta dirección, o alcanzar o apartarse de objetos móviles).

La elección de "experimentar" con jerarquías de mediadores

Una restricción más que debemos atender para nuestra experimentación es la

de los componentes del robot-cognitivo implícito navegante debe tener una adaptación jerárquica. Como ya se mencionó en el capítulo anterior Nuestra fidelidad al paradigma biológico (rana) nos hizo tomar la elección de experimentar con jerarquías.

Arquitectura de nuestros robots navegantes.

Una **arquitectura** se define aquí como la especificación nominal de control mediante listas de funciones. Las funciones se ejecutan condicionalmente, de izquierda a derecha en el orden en el que se encuentran dentro de las listas.

Arquitectura jerárquica general

Un agente reactivo AR se define como un cúadruplo:

Ar := (Sensor Mediator Motor Acción)

donde

Sensor := Sx x lista-sens

Sx := <emisor odométrico en la dirección x >

lista-sens := (S_i S_j ... S_n)

Motor := Motx x lista-mot

Motx := <definición de parámetros motores del esquema motor x>

lista-mot := (Mot_i Mot_j ... Mot_m)

Mediator := Medx x lista-med

Medx := (Perceptor Mediator Acción) x

(Perceptor Motor Acción) x

(Sensor Motor Acción) x (Nil Nil Nil)

lista-med := (M_i M_j...M_n)

Perceptor := Px x lista-percep

Px := <perceptor simple de nombre x que comanda un Sensor e interpreta la información de este y otros sensores >

Px := Sy x lista-sens

lista-percep := (P_i P_j ...P_n)

Acción := Ar x (Acción Am)

Ar := <Modificación del mundo por reubicación del robot>

Am := <Modificación m del resto del mundo>

Notación:

“ := ” debe leerse como “se define como”

“ x ” debe leerse como el “o-exclusivo”

“ < > ” debe leerse como el contenido de las listas y son notas aclaratorias.

<nota: cada uno de los elementos de la arquitectura puede verse además como una función *niládica* con valor nulo excepto Medx o lista-med que entregan siempre un valor binario: 1 si la función produce por lo menos una acción, y 0 de otra manera>.

Control.

CONTROL:=

1. Inicializar los sensores.
2. Buscar en profundidad el primer mediador de la lista anidada de ellos que dé un valor booleano verdadero: (llamar a cada una de las funciones perceptor o sensor de la primera lista del Mediador; y si se cumplen sus condiciones llamar a cada una de las funciones de la lista en él. Retomamos como valor del Mediador un valor booleano cierto).
3. Regresar a 2.

Una versión más detallada de la programación de estos robots se puede ver en el apéndice A de la tesis.

Leyes motoras y leyes de atención perceptuales.

Una ley motora, es el equivalente metafórico de una ecuación diferencial de los modelos de la Dinámica de la Física. En estos una ley puede expresarse como una velocidad puntual que se cumple en un instante del tiempo (dado un campo de fuerzas). Su solución puede ser, por ejemplo, la trayectoria de una masa, una vez que se especifican las condiciones iniciales. La ley de atención perceptual o la de acción motora, son leyes de conducta que se

ejecutan también en un instante del tiempo. Nuestras funciones Px y MoTx son tales leyes.

Las leyes y conductas motoras en nuestros robots.

La salida de cada ley-motora individual, es una re-estipulación de velocidad y de dirección del movimiento del robot. La intervención repetida de estas leyes produce conductas como:

<u>C.Bordea</u>	(Ley MoTbordea) seguir una pared.
<u>C.No-chocar</u>	(Ley MoT_no_chocar) disminuir la velocidad.
<u>C.Seguir-paredes</u>	(MoT_sigue_paredes) moverse paralelamente a una pared.
<u>C.Evita</u>	(Ley MoT_evita) moverse alejándose de una barrera.
<u>C.Tactil</u>	(Ley MoT_tactil) alejar la parte del robot cercana de un objeto.
<u>C.Deriva</u>	(Ley MoT_deriva) moverse en una dirección relativa a una "brújula".

Las leyes de atención perceptiva en nuestros robots.

Cada ley de atención perceptual, está creada para generar solamente la información que es necesaria para una tarea en particular sin la creación de mapa alguno.

Por ejemplo, con los cuatro sensores frontales de NAVEGANTE II "implementamos" una "visión" primitiva para perseguir o apartarse de un objeto móvil sin generar alguna escena del entorno.

Las leyes de acción motora y de atención perceptiva como campos internos del robot.

Una noción muy conocida en matemáticas es la representación de las ecuaciones diferenciales como campos de pendientes. La solución de una ecuación diferencial es la curva trazada "sobre" un grupo de pendientes dada una condición inicial. Análogamente, la trayectoria de una masa puede trazarse en un campo de velocidades puntuales dadas las velocidades iniciales.

La diferencia en cuanto al origen del campo en las velocidades entre una masa en movimiento y el de velocidades de un robot, es esencial, ya que en la primera el campo es un campo externo, mientras que el caso del robot, el campo es generado por el motor del robot en cada punto de su trayectoria (un campo interno).

Este tipo de representación vectorial del campo interno se ha empleado para ilustrar el caso una ley de acción motora (Figura 8.1) y el de una ley de atención perceptiva (Figura 8.2).

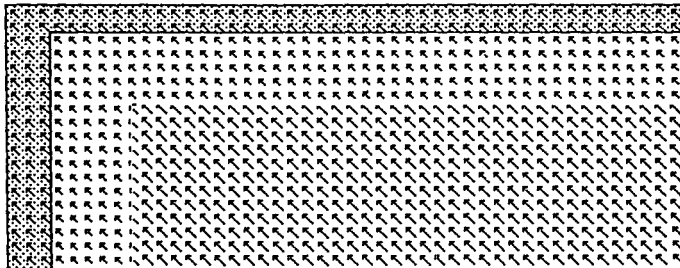


Figura 8.1. Representación como campo vectorial de la ley motora correspondiente al mediador no-choques

Se trata de las dos leyes que constituyen el *mediador no_choques*. Nótese que los vectores de velocidad de la acción motora disminuyen de tamaño en algún lugar del campo. Por otra parte los vectores de atención se dirigen desde los objetos percibidos a la posición del robot.

La figura 8.1 representa del cambio de velocidad que se produce en los motores del robot, para cada punto del *mundo* (el espacio y el marco que lo contiene).

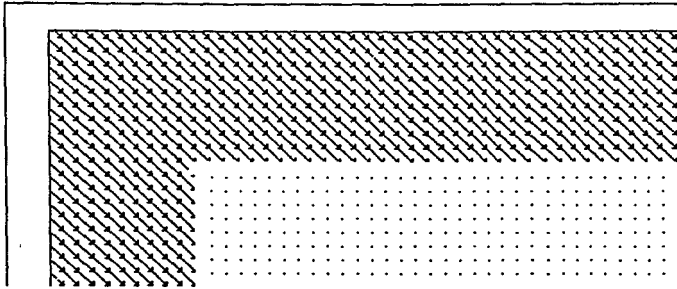


Figura 8.2. Ley de atención perceptiva no-choques

En la figura 8.2, los vectores representan la distancia del marco al punto de contención, medidos por los Odómetros. Los puntos sin flecha indican que el Odómetro está activado fuera de "rango".

El acoplamiento entre leyes de atención perceptual y de acción motora.

Las leyes de atención perceptual y las de acción motora se pueden acoplar mediante un *mediador*. Este acoplamiento tiene la ventaja de permitir que una misma ley perceptual se pueda acoplar con varias leyes motoras (con ahorro de *código*) y además permite formar unidades de conducta más complejas por asociación de mediadores.

Conceptualmente la postulación de mediadores permite establecer una competencia (o cooperación) entre ellos que se puede interpretar como una interacción de *micro-objetivos explícitos*: un *mediador* representa un *micro-objetivo explícito* conductual. El conjunto de todos los *mediadores* interaccionantes que constituyen en robots, es un *Mediador* a su vez, al que se le asigna el *macro-objetivo implícito*: el del *robot*.

El mediador como unidad conductual.

El *mediador* debe considerarse el equivalente a la *unidad conductual* de los etólogos, que les permite realizar *análisis* de conductas complejas. En nosotros la unidad conductual Mediador nos permite efectuar una *síntesis* de ellas.

“Implementación” de movimiento, de sensores y de perceptores en nuestros robots navegantes.

Nuestros robots reactivos se han simulado en lenguaje C en la pantalla de una microcomputadora.

Simulación de movimiento del robot.

Los robots, en nuestra simulación, están representados por los tres vértices de un triángulo en la pantalla de un monitor de computadora (dos puntos extremos y un punto *central*). Los puntos son de color azul en fondo negro): su esqueleto. Cada vez que el robot se *mueve* se borran los tres puntos y se redibujan en la nueva posición del robot. El punto central indica la dirección del robot. El movimiento de los tres puntos debe interpretarse como dos pasos sucesivos del robot, uno con su extremo derecho girado sobre el otro (fijo) y el segundo como un giro del extremo izquierdo, tomando como fijo el nuevo punto derecho. Los ángulos de giro de cada paso están comandados por los controles (Motx, véase la arquitectura en la sección anterior).

Simulación de odometría.

Desde cada punto del esqueleto del robot exploramos el color de la pantalla en distintas trayectorias rectas, buscando un pixel blanco. De detectarse éste, se almacena el número de pixeles explorados sin éxito, como la distancia del punto inicial de exploración (punto de emisión) al objeto blanco encontrado. Esta simulación es equivalente a la medición de distancia del emisor a los objetos. Adicionalmente en algunos casos, medimos el número de pixeles blancos que tiene el objeto encontrado en la trayectoria y este dato se

8 La navegación reactiva

almacena (temporalmente) como el *espesor* del objeto. Los odómetros (o sensores) se clasifican en frontales y laterales por la dirección de las emisiones y su alcance. Dentro de los frontales (que emiten perpendicularmente a la línea que une los puntos extremos) contamos con: el central y los frontales derecho-izquierdo. Dentro de las laterales, que emiten en la dirección de la línea que une a los dos extremos del esqueleto, contamos con los de largo alcance, lateral derecho y lateral izquierdo, y los de corto alcance: táctil derecho y táctil izquierdo. La percepción de un *objeto móvil* está determinada por las distancias y espesores de un objeto en el *campo visual* de los tres sensores frontales. La percepción de *pared* está determinada por su espesor.

Experimentos navegacionales con distintos grados de cognitividad.

El marco en que se *mueven* nuestros robots y los objetos dentro de él son de color blanco. Las líneas horizontales del marco y la de los objetos fijos insertos son interpretados por el robot como paredes, por su espesor. Las líneas verticales del marco son interpretadas por el robot como objetos no-paredes, por su espesor.

En la pantalla del monitor los puntos del esqueleto del robot son azules. En las figuras 8.3 a 8.6 está representada la traza de los puntos extremos del robot en puntos negros en fondo blanco y en las figuras 8.7 y 8.8 se representa la traza del punto central del robot.

En todas las figuras, los marcos y objetos están representados en negro sobre el fondo blanco

La conducta generada por la aplicación repetida de la ley conductual de atención perceptual motora táctil (Mtáctil) por definición es no cognitiva. La figura 8.3 ejemplifica esta conducta en el marco que restringe los movimientos del robot. En esta figura el robot *invade* la pared inferior.

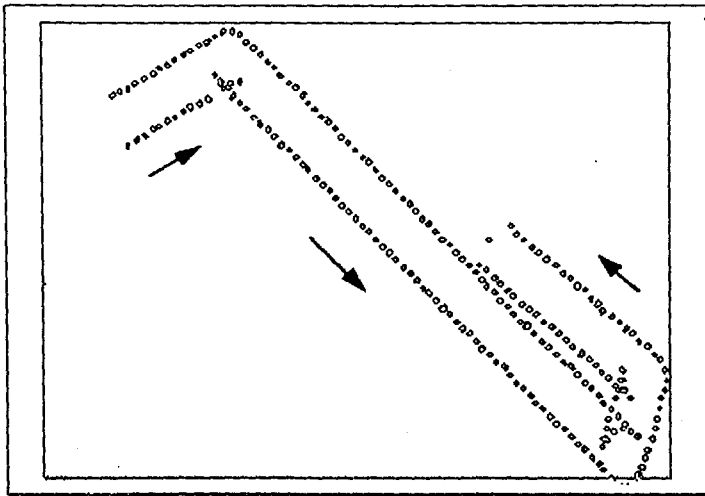


Figura 8.3 Representación en aislamiento de la conducta táctil del robot.

(Mevita>Mdetente) y (Mbordea>Mdetente) están ejemplificadas en las figuras 8.4 y 8.5 respectivamente. Mbordea a su vez jerarquiza (Mno_choques>Msigueparedes)

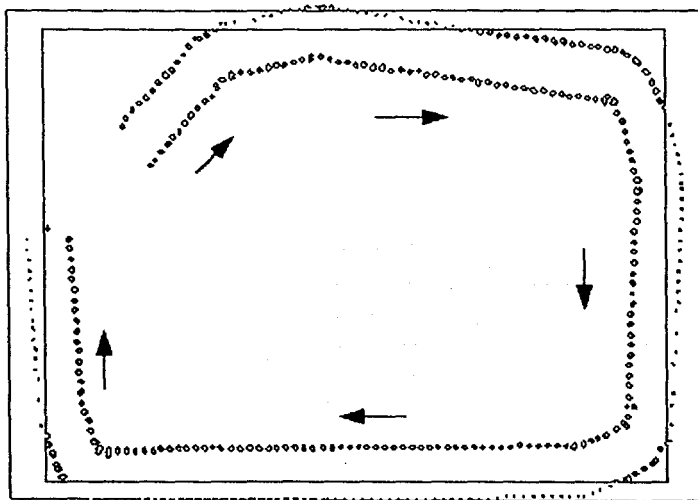


Figura 8.4 Representación de la conducta evita aislada

Las conductas representadas en las figuras 8.3 y 8.4 son inadecuadas. La conducta táctil de la figura 8.3 es inadecuada por definición. La conducta evita de la figura 8.4 porque se invade el marco de contención del robot.

En el caso de la conducta generada por $(M_{evita} > M_{detente})$ el robot invade todos los marcos y en el caso de la conducta $M_{bordea} > M_{detente}$, el robot se detiene frente a la pared derecha.

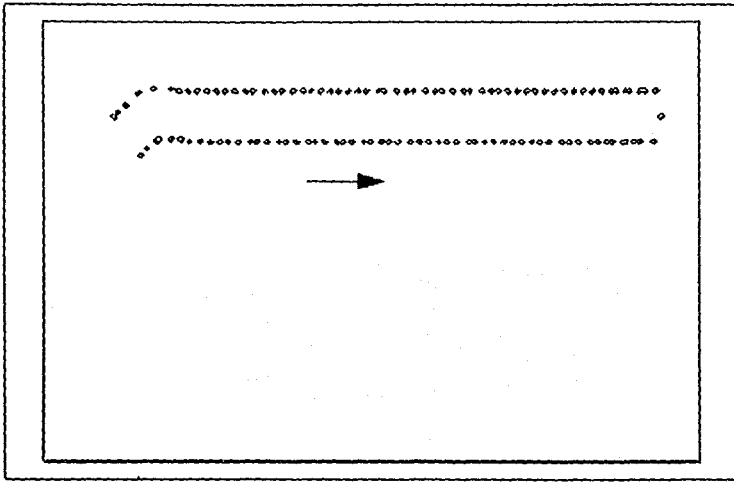


Figura 8.5 representación de la conducta sigue-paredes, resulta inadecuada el robot se detiene frente a una pared derecha.

La figura 8.6 ejemplifica la conducta cognitiva generada por la de la jerarquía de mediadores (en el robot NAVEGANTE I):

Mtactil>Mevita>Mbordea>Mderiva>Mdetente.

NAVEGANTE

I (el robot se detiene), cuando la jerarquía es inversa a la anterior.

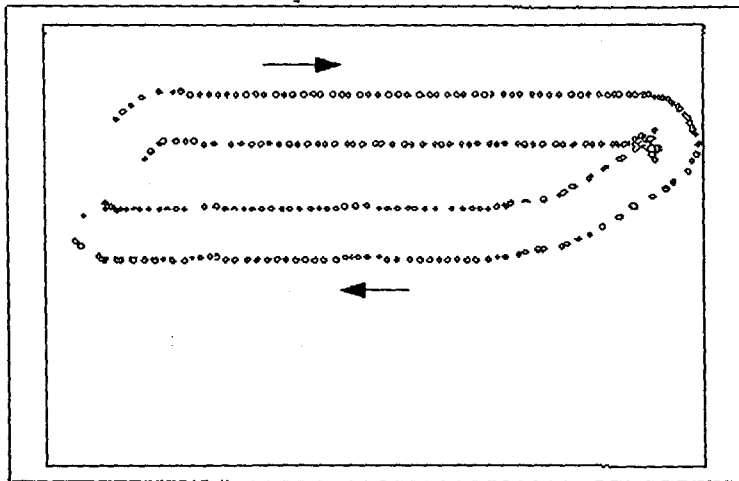


Figura 8.6 Representación de la conducta NAVEGANTE I, adecuada puesto que el robot no se detiene frente a ninguna pared. Esta conducta es el resultado de la intervención jerárquica de varias leyes

La figura 8.7 ilustra la conducta propositiva (direccional) de NAVEGANTE I según el orden jerárquico que se señaló para la figura 8.6. Nótese la habilidad del robot para sortear obstáculos y su tendencia a permanecer en el borde izquierdo del marco, dirección que se ha programado para su *deriva*.

alcanza lo *toma* (la hace desaparecer de la pantalla y el programa pone un nuevo objeto móvil en ella con distinto espesor). Esta ley está constituida por la jerarquía:

(Maproxima>Mtoma).

En la figura 8.8 los dos objetos móviles se mueven pegados al borde izquierdo del marco, en una trayectoria repetida de arriba hacia abajo y de abajo hacia arriba.

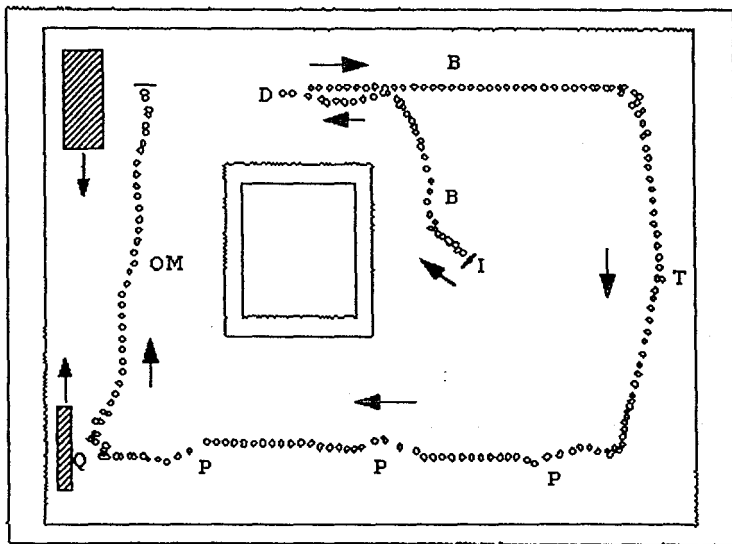


Figura 8.8. Representación del centro del "esqueleto" de NAVEGANTE II. Nótese que el robot tiene una conducta adecuada. Huye de los objetos grandes móviles, D, y sigue P, y captura objetos pequeños móviles Q.

En la figura 8.8 se ilustra una de las posiciones de un objeto-objetivo. En esta figura en

- (i) se inicia el robot.
- (b) se denota la intervención del Mediador Mbordea.
- (d) el robot hace intervenir Malejate al detectar un objeto móvil *peligroso* en el borde izquierdo del marco.
- (t) ocurre la intervención de Mtactil.
- (p) se ha marcado la intervención de Mproxima dentro de Mbusca.

La última p de la izquierda corresponde al momento de la intervención de Mtoma dentro de Mbusca y OM es la trayectoria final del ejemplo.

Principios cognitivos sugeridos por nuestra experimentación

Principios sugeridos por NAVEGANTE I

I. Principio de agrupación motivacional.

Al observar el desempeño del robot se tiene que los **mediadores se agrupan por motivaciones**, por ejemplo: Motivación no-exploratoria y motivación-exploratoria:

(Medente)(Mevita Mtactil Mderiva Mbordea)

II. Principio de jerarquización motivacional.

Si la motivación exploratoria es predominante entonces tenemos que

(Mderiva Mbordea Mevita Mtactil) > (Medente)

Si aplicamos el principio anterior (I) en el primer grupo (suponiendo las motivaciones de conservar la integridad física para el segundo subgrupo y el explorar para el primer subgrupo) entonces se obtiene:

((Mderiva Mbordea) (Mevita Mtactil)) > Medente

III Principio de jerarquización consumatoria.

(ya postulado por [Maes91])

Este principio postula que las conductas consumatorias (de conseguir el objetivo) tienen jerarquía más baja (y también los grupos que las contienen). Entonces, puesto que Mderiva es consumatoria para NAVEGANTEI:

$$((\text{Mevita } \text{Mtactil}) > (\text{Mderiva } \text{Mbordea})) > \text{Mdetente}$$

IV Principio de jerarquización conservacional de la integridad física.

Se postula con este principio que el mediador que protege de una manera más efectiva la integridad del robot tiene mayor jerarquía. Aplicando este principio sobre el primer grupo se obtiene:

$$((\text{Mtactil } \text{Mevita}) > (\text{Mderiva } \text{Mbordea})) > \text{Mdetente}$$

La aplicación del principio III sobre el segundo subgrupo del primer subgrupo da:

$$((\text{Mtactil } \text{Mevita}) > (\text{Mbordea} > \text{Mderiva})) > \text{Mdetente}$$

La aplicación del principio IV sobre el primer subgrupo del primer grupo da:

$$((\text{Mtactil} > \text{Mevita}) > (\text{Mbordea} > \text{Mderiva})) > \text{Mdetente}$$

Aplicación de los principios sugeridos por NAVEGANTE I en NAVEGANTE II

V Principio de agregación sobre estructuras anteriores (ya postulado por [Brooks86])

Si postulamos que NAVEGANTEII es el resultado de la agregación de nuevos mediadores sobre los anteriores, entonces se tiene lo siguiente:

Mbusca Malejate (NAVEGANTEI)

VI Principio de agrupación de antagonicos

Como Mbusca y Malejate son antagonicos por lo tanto:

(Mbusca > Malejate) (NAVEGANTEI)

VII Principio de jerarquización de antagonicos

Los mediadores se jerarquizan por su prioridad para mantener la integridad del robot:

(Malejate > Mbusca) (Navegante)

Aplicando el primer principio I, suponiendo que las motivaciones para intervenir sobre objetos móviles son dominantes tenemos que:

(Malejate > Mbusca) > (NAVEGANTEI) \equiv NAVEGANTEII

Finalmente, la aplicación del principio consumidor en los mediadores de Mbusca produce la jerarquía (Mproxima Mtoma).

La anterior explicación *post-hoc* de las arquitecturas de NAVEGANTEI y de NAVEGANTEII no constituyen en sí un algoritmo de autoprogramación de mediadores, sólo pretende exhibir algunos principios que serían necesarios para la generación de tal algoritmo.

9 Nuestros principios cognitivos y consideraciones finales

Hemos decidido finalizar la tesis con la ubicación de nuestros robots navegantes reactivos en el panorama de los modelos biológicos de batracios que con otros fines han construido otros autores.

Concluimos la tesis con nuestra opinión sobre el valor robótico que tienen los principios postulados por nosotros, y la importancia que tienen ya para nuestro trabajo futuro.



Mapeo de algunas funciones de nuestros robots navegantes en la neuroanatomía de los batracios

Hemos creído que el lugar más apropiado de la tesis para comparar nuestro trabajo con el de otros autores es el final, aún cuando nuestros robots no tienen la intención de reproducir la conducta de ningún batracio sino la de suscitar principios cognitivos para el diseño robótico

El modelo de [Lara88]

La figura 9.1 es esencialmente la misma que la figura que mencionamos al inicio de la tesis, pero en ella hemos suprimido las interacciones entre los núcleos, y en su lugar hemos sobrepuesto algunas de las funciones importantes que arquitecturan nuestros robots navegantes. El único mediador que pudimos mapear sobre la figura original de Lara es el táctil, sobre la región conjunta tegmentum y médula, constituido por su parte perceptual Ptactil y su parte de acción Mottactil. Su entrada proviene del mundo exterior y va hacia los motores del robot (los músculos del animal).

Nuestro perceptor de objeto fijo podría mapearse en el tálamo anterior, con su entrada proveniente de los tres odómetros frontales, equivalentes a una retina primitiva. La salida del perceptor de objeto fijo, en nuestra programación, es una función con valor binario que está al *servicio* de distintas leyes de acción mapeadas conjuntamente en el tegmentum y médula. Por definición, el valor 0 significa un objeto móvil, por lo que éste lo hemos mapeado sobre la región del tectum para establecer la correspondencia deseada.

El mapeo de la función perceptora de objeto pequeño, la hemos mapeado sobre el pretectum, y como esta función también entrega una salida binaria. La salida 1 activaría varias leyes motoras en el tegmentum y médula, y puesto que por definición la salida significa la percepción de un objeto grande, este valor lo hemos mapeado sobre el tectum. La entrada del perceptor objeto pequeño es la misma que la de objeto fijo: la de los tres odómetros frontales. Nótese que no existe en nuestros robots, al igual que en

los batracios, una interacción de percepciones, como corresponde a la arquitectura robótica reactiva. Los mediadores, unidades elementales de control, no son mapeables, excepto el que ya mencionamos, en una topología neuroanatómica como la reportada por Lara.

Finalmente, de la misma manera que Lara asigna las motivaciones al telencéfalo, nosotros mapeamos nuestros principios cognitivos a esta estructura. Nótese que implícitamente planteamos, con la discusión anterior (del mapeo), la tesis de arquitecturas equivalentes entre la nuestra, por ejemplo, y la biológica. La nuestra prácticamente *sin representaciones* y la biológica con las representaciones propuestas que se considera que existen en ella, véase el capítulo 1.

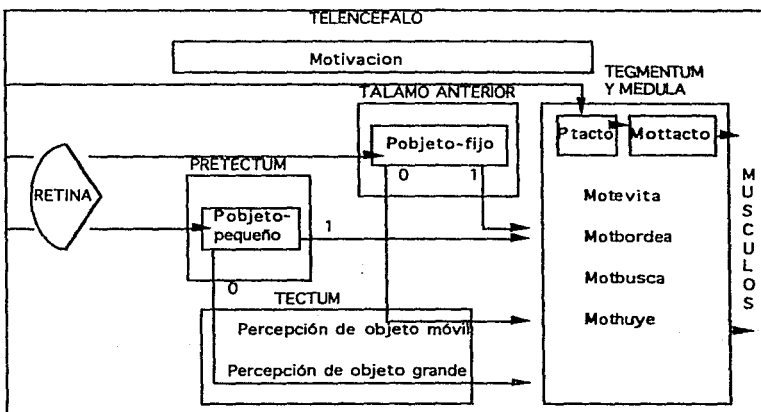


Figura 9.1 Nuestra arquitectura sobrepuesta sobre el modelo de Lara de la neuroanatomía de la rana.

Otra definición fundamental entre estas arquitecturas equivalentes es la imposibilidad de mapear el control de una en otra. Sin embargo, una analogía persiste: la jerarquización del telencéfalo y otros núcleos.

El modelo de Arbib¹

La figura 9.2 presenta el esquema-modelo de Arbib de la conducta de depredación de la rana. Nótese que esencialmente se trata de un modelo representacional de escenas, que se activan entre sí. Nótese además, que se postulan efectos inhibitorios de acciones como en el modelo clásico de Lorenz (véase la figura 1.1)

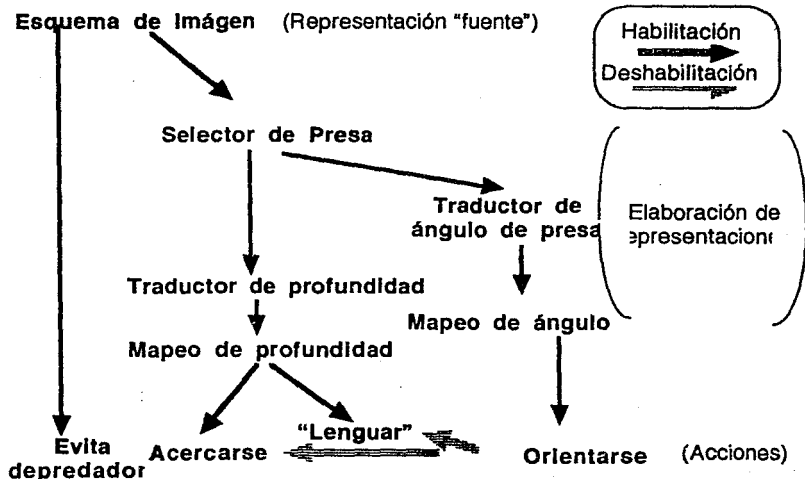


Figura 9.2 Modelo de esquema de Arbib () para un agente depredador que puede huir de otro depredador.

Nótese la existencia de una representación "fuente" que se utiliza para elaborar las representaciones pertinentes a las acciones de tomar. Nótese también, que el esquema corresponde en parte al modelo clásico de la etología de Lorenz en cuanto a la inhibición de acciones se refiere.

La Figura 9.3 hace ver claramente que el mapeo de nuestro control

¹ op. cit. Arbib

jerárquico sobre el esquema de depredación de Arbib no es posible porque:

- a) No contamos con un sistema representacional y;
- b) No nos basamos en el modelo etológico clásico de Lorenz de inhibición recíproca de conductas.

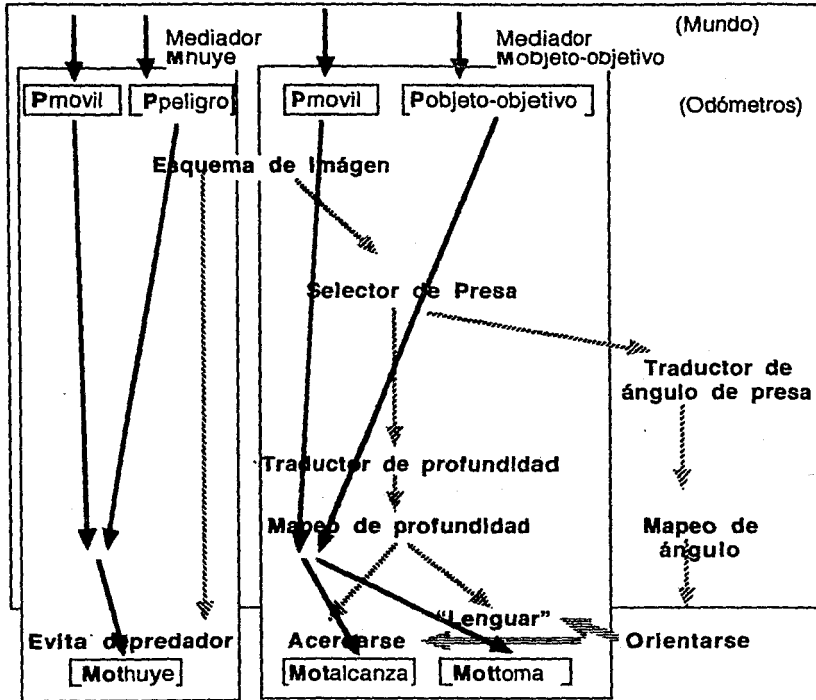
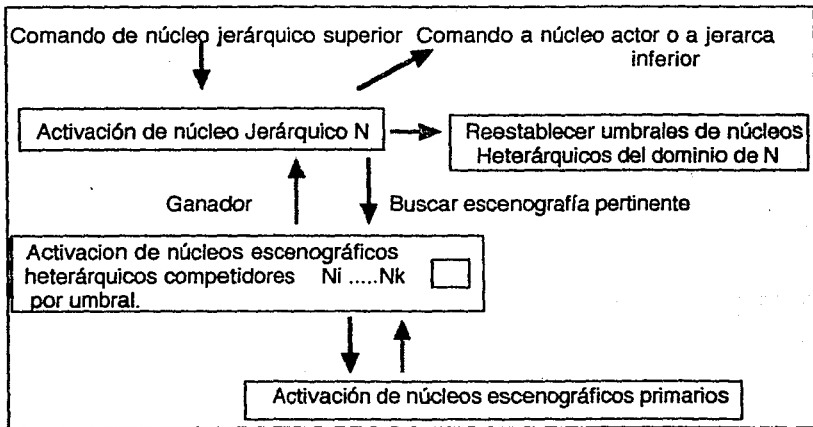


Figura 9.3 Mapeo parcial de nuestro robot Navegante III (insertos) sobre el esquema de Arbib (negritas).

En nuestro robot está claramente ausente además de la representación de objetos, la de efectos por la inhibición de acciones. El original **Mfijo** se ha cambiado por **Mmóvil** para facilitar el mapeo. El robot controla las acciones por jerarquía de mediadores (**Mhuye** > **Mobjeto-objetivo**).

Modelo jerárquico-heterárquico de Negrete y Lara.

[Negrete-Lara86] proponen un modelo de sociedad de núcleos al estilo Minsky. Esencialmente se trata de núcleos representacionales que capturan escenas del mundo de la rana (escenográficos: células TH-3, columnas tectales, etc.) y que de una manera generalizada se puede esquematizar en la figura 9.4



En este un modelo jerárquico-heterárquico de núcleos de batracio, cada jerarca tiene un dominio de núcleos heterárquicos que designan ganador por competencia a base de umbrales (figura 8-4). El enfoque esta centrado en

sugiere un paradigma interesante para ser desarrollado: un sistema reactivo JERARQUICO-HETERARQUICO.

El poder predictivo de nuestros principios

A manera de recordatorio reproducimos aquí los enunciados de nuestros principios:

- I Principio de agrupación motivacional.**
- II Principio de jerarquización motivacional.**
- III Principio de jerarquización consumatoria. (postulado por Maes(12))**
- IV Principio de jerarquización conservacional de la integridad física.**
- V Principio de agregación sobre estructuras anteriores**
- VI Principio de agrupación de antagonicos**
- VII Principio de jerarquización de antagonicos**

De acuerdo con [Dennett89], los principios cognitivos encontrados deben tener cierto poder predictivo conductual para que sean válidos; además de que deben proveer una guía para el diseño:

Curiosamente la adaptabilidad de las conductas es una característica que nuestro principio II predice, aunque no es de extrañar tal predicción puesto que aquélla es una característica fundamental de los robots reactivos. La robustez vista como adaptabilidad es también otra de las propiedades inherentes a nuestra arquitectura y a la naturaleza reactiva de ella.

La reflexibilidad (de reflejo) es también una característica etológica fundamental, particularmente predecible por nuestro principio IV.

El obvio principio conductual de exclusión de conductas antagonicas puede también ser predicho por nuestro principio IV.

Puesto que nuestros principios están postulados en términos de motivaciones, éstos también deberían poder predecir algunas de las características conductuales típicas de las llamadas conductas motivadas (véase el capítulo 1)

9 Consideraciones finales

- 1) El principio I predice el agrupamiento de componentes conductuales en el tiempo y el principio III su secuenciación.
- 2) El principio V predice la direccionalidad a objetivos cuando por ejemplo, las nuevas jerarquías agregadas tienen en sí mismas las motivaciones que explican los nuevos objetivos.
- 3) El principio V también predice la espontaneidad conductual ya que los robots constantemente atienden a los objetivos de los mediadores jerárquicamente superiores.
- 4) No son predecibles los cambios de responsividad, su persistencia de objetivos, ni el aprendizaje asociativo.
- 5) nuestros principios no predicen tampoco algún tipo de plasticidad conductual.

Surgen de nuestro trabajo tres preguntas importantes relativas a los principios cognitivos por nosotros propuestos--sin referencia alguna a su relevancia biológica:

- a) ¿Son estos principios guías para el diseño de robots que con la adición de mediadores jerárquicamente superiores exhiban conductas de orden superior ?
- b) ¿Son estos principios también guía para el diseño de algoritmos de autocontrol *jerárquico-heterárquico* ?
- c) ¿Nos ofrecen estos principios la cara opuesta a la de máquina de la dialéctica mente-cuerpo ?

Creemos que todas las preguntas deben contestarse afirmativamente pero es la primera la que está determinando ya nuestro trabajo de investigación en un futuro inmediato.

Referencias

- [AgreP-ChapmanD87] Agre, Philip E y Chapman David; Pengi: an implementation of a theory of activity, **AAAI87**: 268-272
- [ArbibMA84] Arbib; M.A. *Rana* computatrix: progress report, 1984. **Federation Proc.** 44: 2956-2958; 1985
- [Beer-Chiel91] Beer, Randall D. y Chiel H.J; The neural basis of behavioral choice in an artificial insect; **From animals to animats: Proceedings of the First International Conference of Simulation of adaptive behavior**; 1991
- [Brachman-Levesque85] Brachman Ronald y Levesque Héctor; **Readings in knowledge representation**; Morgan Kauffman, 1985
- [Brooks86] Brooks, Rodney A.; A robust layered control system for a mobil robot; **IEEE J. on Robotics and Automation**; RA-2; 1986: 14-23
- [Brooks89] Brooks, Rodney A.; A robot that walks: Emergent behavior from a carefully evolved network; **Neural Computation**; 1(2), 1989: 252-262
- [Brooks90] Brooks, Rodney A., "Elephants don't play chess". En **Designing Autonomous Agents: theory and practice from biology to engineering and back**, Pattie Maes (ed). MIT Press, 1990. pp. 3-15.
- [Brooks90a] Brooks, Rodney A.; Intelligence without representation; **Artificial Intelligence**, 1990

Referencias

- [CervantesF-LaraR-ArbibMA85] Cervantes, F; Lara, R.; Arbib, M.A. A neural model of interaction subserving prey-predator discrimination and size preference in anurans. **J. Theor. Biol.** 113: 117-152;1985
- [Connell87] Connell, Jonathan; A creature building with the subsumption architecture; **IJCAI-87**; Milán 1987: 1124-1126
- [Connell88] Connell, Jonathan; A behavior-based arm controller. **MIT AI memo 1025** (june 1988)
- [CovirgaruA-LindsayR91] Covirgaru, Arie A. y Lindsay, Robert K. Autonomous deterministic systems. **AI Magazine**, Otoño 1991.
- [Cruse91] Cruse, Holk; Coordination of leg movements in walking animals; **From animals to animats: Proceedings of the First International Conference of Simulation of adaptative behavior**;1991:105-119
- [Chapman87] Chapman David; Planning for conjunctive goals; **Artificial Intelligence**; 32(3); 1987: 333-377
- [DennettD78] Dennett, D. Why not the whole iguana. (comentarios sobre Pylyshyn 1978). En **Behavioural and Brain Sciences**,1978; pp 103-104.
- [DennettD89] Dennett D. **La actitud intencional**. Gedisa. 1989.
- [DewdneyAK84] Dewdney, A.K., **The Planiverse: computer contact with a two-dimensional world**, Poseidon Press, 1984

Referencias

- [Erceau-Ferber91] Erceau, J.E. y FerberJ; **L'Intelligence Artificielle distribuée; La Recherche**; num 233 (juin); 22:750-758; 1991
- [Flynn87] Flynn, Anita; Gnat robots (and how they will change robotics), **IEEE Micro Robots and Teleoperators Workshop**; 1987
- [Flynn88] Flynn, Anita; Gnat robots a low-intelligence and low-cost approach **IEEE Solid-State Sensor and Actuator Workshop**; 1988
- [Flynn-Brooks89] Flynn, Anita, Brooks, R.A. y Tavtow L.S. Twilight zones and conerstones: a gnat robot double feature; **MIT AI memo 1126 (july 1989)**
- [GeorgeffM91] Georgeff, Michael P., Planning. En **Autonomous Mobile Robots: Control, planning and architecture**; Iyengar,S; Elfes, A (eds); IEEE Computer Society Press, vol.2; 1991:359-400
- [HorswillI-BrooksR88] Horswill, Ian D. y Brooks, Rodney A.; Situated vision in a dynamic world: chasing objects, **AAAI88**: 796-800
- [Johnson87] Jonhson, M. **The body in the mind**. University of Chicago Press, 1987.
- [KaelblingLP87] Kaelbling, Leslie P., "An architecture for reactive systems". En Michael P. Georgeff y Amy L. Lansky (eds), **Reasoning about actions and Plans**, Morgan Kauffman, 1987; pp. 395-410.

Referencias

- [LaraR89] Lara y Zavala, Rolando. **Cibernética del cerebro**. CECSA, 1989
- [LloydD89] Lloyd, Dan, **Simple Minds**. (MIT Press, 1989)
- [McCarthy79] McCarthy, John; Adscribing mental qualities to machines; **SRI Technical Report** memo 326; Stanford AI Lab; 1979
- [Maes91] Maes, Pattie; A bottom-up mechanism for behavior selection in an artificial creatureen **From Animals to Animats**. Proceedings of the First International Conference on Simulation of Adaptative Behavior. Meyer, J.A. y Wilson, W. eds. MIT Press. 1991: 238-246
- [Mataric89] Mataric, Maja; Qualitative sonar based enviroment learning for mobile robots; **SPIE Mobile Robots**; 1989
- [NegreteJ91] Negrete, José, **Imágenes preinteligentes, Información Científica y Tecnológica**, CONACYT, vol 13, no. 177, Mexico, 1991
- [NegreteJ-LaraR82] Negrete, José y Lara, Rolando, Visual Nuclei as a society involved in visuomotor coordination in toads, en **Proceedings of the Workshop on Visuomotor in Frogs and Toads: Theory and Experiments**, Mexico, 1982.
- [NewellA-SimonHA76] Newell, Allen y Simon, Herbert A., "Computer Science as Empirical Inquiry: symbols and search",

Referencias

- Comm. ACM**, 19, 1976; pp. 113-236
- [Nilsson84] Nilsson, J. Nils; Shakey the robot. **SRI Technical Note 323**, Stanford AI Center; 1984
- [SchnepfU91] Schnep U. "Robot Ethology: A proposal for the research into Intelligent Autonomous Systems", en **From Animals to Animats**. Proceedings of the First International Conference on Simulation of Adaptative Behavior. Meyer, J.A. y Wilson, W. eds. MIT Press. 1991; pp 465-474
- [SearleJ80] Searle, John, **Actos del habla: ensayos de filosofía del lenguaje**, Ediciones Cátedra, 1980
- [Shoham90] Shoham, Yoav; **AGENT0 An agent-oriented programming**; **SRI Technical Report** memo; Stanford AI Center, 1990
- [Smith80] Smith, R.G.; "The contract-net protocol: high level communication on control in an distributed problem solver"; **IEEE Transactions on Computers**; 29(2):1104-1113; 1980
- [SteelsL90] Steels, Luc; Cooperation between agents through self-organizing; **Proceedings of the Workshop of multiagent cooperation**. Cambridge U.K., North Holland, 1990
- [Sussman75] Sussman J.; **A computer model of skill acquisition**; Elsevier, 1975
- [Varela90] Varela, F. **Conocer**. Gedisa. 1990.

Referencias

[Walter50]

Walter, W. Grey.; An imitation of life; **Scientific American**; 182(5): 42-45; may; 1950

[Wiener61]

Wiener, Norbert; **Cybernetics or Control and Communication in the animal and the machine**; The MIT Press, 1961