Arbor



# Inteligencia artificial, emoción y neurociencia

*J. Mira*

Arbor CLXII, 640 (Abril 1999), 473-506 pp.

*Presentamos en este trabajo un conjunto de interrelaciones entre Neurociencia y Computación desde una perspectiva metodológica común, basada en la distinción de tres niveles (fisiológico, de los símbolos y de conocimiento) y dos dominios de descripción (el propio de cada nivel y el del observador externo). Se explora la potencial validez del paradigma computacional para ayudar a la explicación del funcionamiento del sistema nervioso. De forma complementaria, se razona también sobre la validez de la Neurociencia como fuente de inspiración en el campo de la Inteligencia Artificial.*
*Finalmente, se analizan algunos de los términos básicos que, procedentes de la esfera semántica de lo emocional, han invadido recientemente el campo de la robótica y la Inteligencia Artificial.*

## 1. Introducción

Hasta hace poco la inteligencia artificial (IA), en sus dos perspectivas básicas (la simbólica y la conexionista), ha estado dominada por su intento de hacer computable una parte importante del conocimiento humano no analítico. Es decir, pretendía modelar primero y construir después programas de computador que emularan aspectos no triviales del conocimiento humano en tareas científico-técnicas, tales como la solución de problemas, el diagnóstico, el control, el diseño o la enseñanza de determinadas materias. Esta perspectiva aplicada de la IA ha dado lugar a los llamados «Sistemas Expertos» o, de forma más general, a los «Sistemas Basados en el Conocimiento» (18).

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

Paralelamente, desde sus orígenes neurocibernéticos en torno a 1943, con los trabajos de Wiener (37,31), W.S. McCulloch (13,14) y K. Craik (4), la IA ha perseguido el viejo sueño griego de mecanizar los procesos del pensamiento, intentando comprender, modelar y simular los procesos cognitivos característicos de nuestro sistema nervioso: las distintas modalidades sensoriales (visión, audición, tacto, ...), la integración plurisensorial, la memoria, el aprendizaje, el razonamiento y el lenguaje natural. Cuando añadimos sensores y efectores a un programa que contiene el modelo de lo que sabemos sobre cómo vemos decimos que tenemos un sistema de *visión artificial*. Y lo mismo con el resto de los procesos cognitivos.

Por decirlo de forma resumida, hasta hace poco la IA sólo se había interesado por intentar comprender y mimetizar, al menos parcialmente, los aspectos cognitivos de la conducta inteligente. Sin embargo, recientemente, el campo de la computación en general y la IA en particular están siendo invadidos por términos antropomorfos procedentes de la esfera emocional (1,8,9), pasando de la búsqueda del «pensamiento artificial» a la del «sentimiento artificial». La robótica autónoma, los brillantes desarrollos en técnicas multimedia y de realidad virtual y el empuje de algunos apartados de la llamada «vida artificial» han potenciado esta introducción del vocabulario del mundo de la emoción al campo de la computación. Ahora es frecuente oír hablar de «agentes benévolos», «intenciones y propósitos de una máquina», «motivaciones de un programa», etc... En mi opinión, esta tendencia nos distrae en el verdadero camino de interrelación entre Neurociencia y Computación. Más adelante dedicaremos un apartado a analizar las razones en las que se basa esta opinión pero primero dedicaremos una parte importante del trabajo a intentar presentar de la forma más clara posible lo que considero que son las bases metodológicas para una relación fructífera entre Neurociencia y Computación, en la tarea común de comprender el funcionamiento del Sistema Nervioso (SN) modelando computacionalmente los resultados experimentales y usando esos modelos en predicción y en la sugerencia de nuevos experimentos dentro de grupos de trabajo de carácter interdisciplinario. Este camino es, probablemente, menos espectacular que el usual pero, seguramente, más serio y eficiente a largo plazo.

El resto del trabajo está estructurado en los siguientes apartados: 2. Fundamentos del paradigma computacional. 3. Aspectos Metodológicos: Niveles y Dominios de descripción comunes al SN y a la Computación. 4. De lo natural a lo artificial (cómo puede ayudar la Neurociencia a la Computación y la Ingeniería). 5. De lo artificial a lo natural (cómo pueden ayudar la Física, la Teoría de Sistemas y la Computación a la comprensión del SN). 6. La Emoción en la IA: ¿Existe

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

la Máquina Emocional? 7. Los propósitos en la IA: ¿Existe la Máquina Intencional? 8. Reflexiones finales.

## 2. Fundamentos del paradigma computacional

### *2.1. Aspectos Generales*

La idea que subyace a todo el movimiento interdisciplinar relacionado con el modelado computacional del SN es que:

*Los seres vivos y las máquinas pueden comprenderse usando la misma metodología experimental, los mismo principios de análisis, los mismos esquemas organizacionales y estructurales y las mismas herramientas formales y computacionales.*

La metodología experimental no es otra que la de la física (en su doble aspecto correlacional y teórico) que nos lleva a construir modelos computables de un segmento fenomenológico. Para que el modelo sea computable ha de cumplir, al menos, las siguientes condiciones.

*a)* Ser *claro, completo, preciso* e *inequívoco*. Es decir, estar suficientemente especificado el nivel anatómico y el fisiológico (26,7) para que nos permita llegar a la formulación lógico-analítica mediante un conjunto de variables de entrada y salida y un conjunto de reglas de transformación computables. En el proceso de construcción del modelo es indispensable especificar las tablas de semántica de la precodificación para dar después significados del mismo nivel en la interpretación de los resultados (24,18).

*b)* Proponer un conjunto de *hipótesis* para correlacionar estructura, función y comportamiento observable y *comprobar* la *completitud* y la *consistencia* interna del modelo (cierre a estructura y organización del nivel fenomenológico considerado).

*c)* Estar *bien fundamentado biológicamente*. Es decir, que las suposiciones sobre las variables y sus relaciones en el modelo no sean arbitrarias respecto a sus correspondientes experimentales.

*d)* Intentar usar pocas hipótesis con gran capacidad de explicación. El modelo debe ser *sintético* y *predictivo*.

*e)* Validarlo y evaluarlo con *datos reales*, procedentes del experimento y usarlo para sugerir nuevos experimentos.

Las herramientas *formales* y *computacionales* también son comunes a seres vivos y máquinas. No tenemos otras (27). Son las propias de la lógica (combinacional y secuencial) y la matemática (ecuaciones integro-diferenciales, cálculo de probabilidades, transformaciones integrales tipo Laplace y Fourier, teoría de circuitos,...) junto con los

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

operadores lógico-relacionales que ofrece un lenguaje de programación de alto nivel para el que exista un compilador. Todo lo que no pueda ser modelado usando estos operadores (sumas, productos, derivadas, integrales y condicionales tipo «*if ... then ...*») no pasará la barrera de un compilador y, por consiguiente, no será computacional en sentido estricto. Podrá, eso sí, ser descrito como conocimiento adicional *inyectado* a los resultados de un programa por un observador externo a la máquina.

Vamos a mencionar ahora alguno de los principios organizacionales y estructurales que, en gran medida, también son comunes a seres vivos y máquinas:

- *Comunicación* (conceptos de señal, mensaje, información, código, canal y ruido).
- *Cálculo* (conceptos de representación fisiológica de símbolos, transformación y almacenamiento y producción de respuestas resultado de esos datos y sus transformaciones, de acuerdo con ciertos contenidos de una «memoria»).
- *Control* (esquemas de realimentación negativa, variables consigna, homeóstasis, lazos múltiples, estabilidad, lazos sensoriomotores).
- *Oscilación* (esquemas de realimentación positiva, temporizadores, relojes biológicos, oscilaciones sincronizadas y cooperativas).
- *Plasticidad* (procesos de maduración, memoria, aprendizaje, ajuste de parámetros, autoprogramación, acoplo estructural, ...).
- *Autoorganización y cooperatividad* (tolerancia a fallos, dinámicas no lineales, factores de escala).

Evidentemente, hay otros principios organizacionales y estructurales que, en el estado actual del conocimiento, deben considerarse como característicos de lo vivo (autopoyesis, los procesos genéticos y evolutivos, la conciencia, la conducta intencional y muchas de las entidades de la esfera emotiva). ¿Qué sentido tendría afirmar que una máquina ama, tiene miedo o sufre? No debemos olvidar la naturaleza esencialmente distinta del soporte físico de las máquinas que calculan (un cristal de silicio) y de los seres vivos (el tejido biológico), junto con la reflexión de Maturana de que «todo conocer depende de la estructura que conoce».

### *2.2. El modelo General de Computación en un Nivel*

En computación usamos el concepto de nivel en dos sentidos: (1) Nivel *fenomenológico* y (2) Nivel de *descripción* (físico, simbólico o de conocimiento).

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

El primer sentido (nivel fenomenológico) es el usual en biología y en física. Focaliza y acota el conjunto de resultados experimentales que se pretenden explicar y define su granularidad. Así, en biología se distingue el nivel físico-químico subcelular, el celular (bioquímico y eléctrico) el orgánico y el de comportamiento global. En Electrónica se distingue entre la electrónica física (teoría de bandas, fenómenos de transporte y recombinación) las estructuras con discontinuidades físicas y eléctricas (uniones), los dispositivos considerados como elementos de circuitos y las funciones de síntesis (amplificación, oscilación, filtrado, ...). Son los niveles de *análisis.*

El segundo concepto de nivel (el *de descripción*) fue introducido en la computación por Allen Newell (28) y David Marr (11). De este concepto de nivel y de las propuestas de Marr («teoría computacional, representación y algoritmo e implementación») y de Newell («nivel de conocimiento, nivel simbólico y nivel físico»), hablaremos en el apartado siguiente. Aquí sólo los mencionamos para dejar claro que en ambos sentidos se tiene que interpretar el concepto de «*modelo general de computación en un nivel*». Es decir, este modelo estructural, afirma la ciencia de la computación, debe ser válido para todo nivel fenomenológico (dominio de conocimiento que queremos modelar) y para todo nivel de descripción de cualquiera de esos dominios de conocimiento.

El modelo general de computación *(18,17) afirma que,* toda la fenomenología de un nivel se puede describir mediante la interacción medio/sistema, tal como se muestra en la figura 1, donde el medio es a su vez otro sistema que puede ser descrito de la misma forma. De hecho, cada partición medio/sistema *define un compartimento de un nivel,* específica un conjunto de señales del medio (que entiende el sistema) y especifica también el conjunto de respuestas del sistema (que entiende el medio). Es decir, cada partición medio-sistema dentro de un nivel queda caracterizada por un lenguaje formal común con el que se describe la interacción medio-sistema (la dinámica de las señales que intercambian).

Por convenio, llamamos *medio* a lo que estimula y *sistema* a lo que responde y el comportamiento del sistema se describe en términos de un conjunto de variables de *entrada* $X = \{x_i(t)\}$ que deben ser medibles, un conjunto de *variables de salida*, $Y = \{y_j(t)\}$ que también deben ser medibles, y un conjunto de *reglas de transformación:*

$$R = \{f_{ij}^{k}(t), g_{ij}^{k}(t)\}$$

que, de forma inequívoca, realizan procesos de cálculo de naturaleza analítica y/o lógico-relacional, sobre las variables de entrada y los con-

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

tenidos de memoria, $M = \{m_n(t)\}$, para generar los valores de las variables de salida:

$$y_j(t+\Delta t)=f^k_{ij}\,[x_i\,(t),m_k\,(t)]$$

y modificar los contenidos de la memoria,

$$m_k(t+\Delta\,t) = g^{\,k}_{ij}\,[x_i\,(t),m_k\,(t)]$$

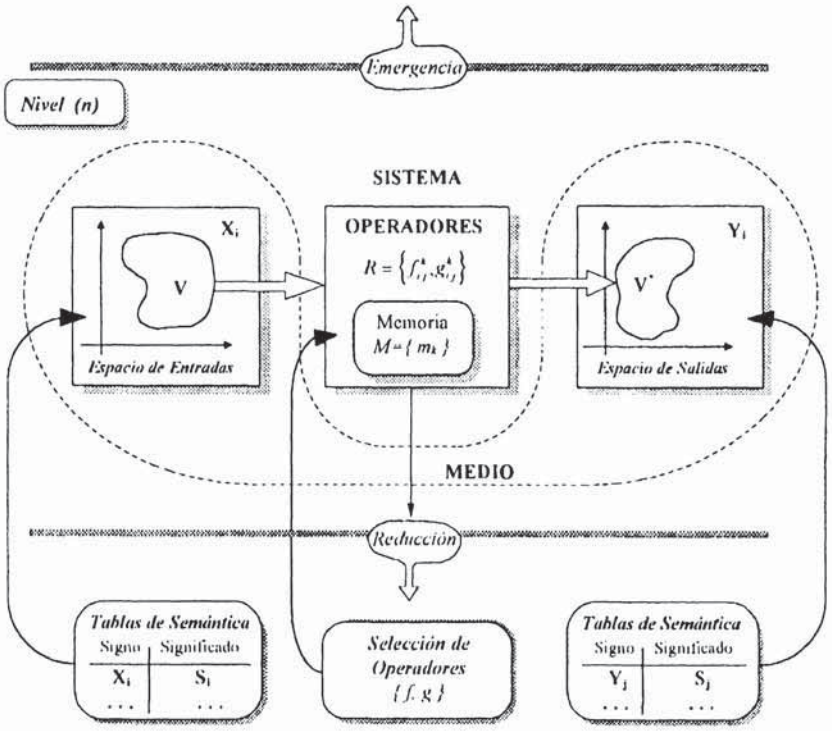
Todo modelo computable en un nivel puede entonces describirse en términos de un conjunto de *señales* (variables $x_i$, $y_j$ y $m_k$) que representan la información (los datos) y un conjunto de reglas (operadores) $f^k_{ij}$ y $g^{\,k}_{ij}$ que especifican de forma «*clara, precisa, completa* e *inequívoca*» los procesos analíticos o lógicos-relacionales que se usan para cualquier secuencia de representaciones de entrada $\{x_i(t),\ x_i(t+\Delta t),\ x_i(t+2\Delta t),\ \ldots\}$ en la correspondiente secuencia de representaciones de salida $\{y_j(t),\ y_j(t+\Delta t),\ y_j(t+2\Delta t),\ \ldots\}$, sin ninguna conexión causal con el *significado* de las variables. Es decir, en principio las variables $\{x_i\}$ e $\{y_j\}$ podrían interpretarse como *magnitudes físicas* (presión, volumen, temperatura, potenciales eléctricos, corrientes, ...) que sirvieran de *soporte material* de la computación. Estaríamos entonces en un *nivel físico*, en el hardware de la computación, en la electrónica digital y la arquitectura de computadores donde las variables $x_i(t)$ e $y_j(t)$ son señales binarias con sólo dos valores posibles (0 ó 5 voltios, por ejemplo), asociados a dos estados lógicos («0» y «1»).

Obsérvese sin embargo que no hay nada en el modelo que nos obligue a esta interpretación. Es decir, las reglas $f^k_{ij}$ y $g^{\,k}_{ij}$ que enlazan los espacios de representación son independientes de la semántica de otros niveles y para conseguir sus resultados formales no necesitan hacer referencia alguna a los significados de las variables. Esto implica que, si en vez de hablar de variables físicas de entrada y salida, hablamos de *espacios de representación* de las entradas y las salidas, podríamos estar hablando de cualquier otro *nivel*. Por ejemplo, del nivel de los símbolos, o del lenguaje natural. O de un sistema neurofisiológico tal como un contacto sináptico, una neurona o una red neuronal del inhibición lateral, por ejemplo. La clave está en las *tablas de semántica* que usemos para describir el *significado* de esas variables y, consecuentemente, el de los *procesos* representados por los operadores que las enlazan. Estos significados siempre tienen que definirse en dos niveles, que a su vez pertenecen a dos dominios de descripción diferentes:

a) En el *dominio propio* del nivel (señales o símbolos), donde hay causalidad y semántica intrínseca al nivel.

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

FIGURA 1. Modelo computacional en un nivel. Los espacios de entrada y salida son en general, espacios de representación con tablas de semántica dependientes del nivel y del conocimiento que se quiere modelar




b) En el *dominio del observador* externo y en el nivel de conocimiento, en el que se puede realizar la descripción en dos tipos de situaciones:
   b.1) Cuando hablo del nivel propio, con sus leyes de causalidad inmutables, asociadas a la estructura (las mismas del apartado a).
   b.2) Cuando asigno significados del dominio cuyo conocimiento estoy modelando y, por consiguiente, puedo usar códigos arbitrarios.

Para el lenguaje de modelado computacional propio de cada nivel se generan y reconocen un conjunto de acoplamientos *(X,S),* donde *X* es la representación física o formal de las señales que acepta el nivel y *S* es la interpretación semántica asignada a *X* por las reglas del lenguaje característico de ese nivel. Los pares signo-significados $[(x_i,S_i),(y_j,S_j)]$ y las funciones $f_{ij}^{k}$ y $g_{ij}^{k}$ caracterizan el modelo en un nivel. Cuando estos pares signo-significado se refieren a los lenguajes del nivel físico o del nivel de los símbolos (los programas), sus significados propios están *predefinidos* y son absolutamente rígidos. En cambio, cuando estamos describiendo o interpretando el modelo

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

a nivel de conocimiento, estos pares (*signo, significado*) pertenecen al lenguaje natural y conllevan mucha más riqueza semántica.

Podemos resumir ahora el proceso del modelado computacional:

1. Selección del *nivel* y del *compartimento* a modelar dentro de ese nivel, distinguiendo entre *medio* y *sistema*.
2. Descripción de los *espacios de entrada y salida* a nivel fisiológico y a nivel de conocimiento, con las adecuadas *tablas de semántica* que establecen de forma inequívoca las correspondencias entre los distintos *significados* (fisiológico y formal) de los mismos *signos*.
3. Selección y conexión de *operadores* de acuerdo con las hipótesis del modelo y obtención de resultados formales. Aquí no hay ninguna referencia a significados.
4. Interpretación de los resultados usando las tablas de semántica definidas en 2. Esta interpretación siempre se realiza a nivel de conocimiento y en el dominio del observador.

Es necesario señalar ya aquí algunos de los problemas que presenta el paradigma computacional en su propósito de comprender el funcionamiento del sistema nervioso: (1) la limitación de los operadores, (2) la limitación en los significados, (3) los errores en el proceso de teorización a partir del modelo.

Veamos primero la limitación de los operadores. Gran parte de lo que se puede obtener de un modelo computacional, y todo lo que no se puede obtener del mismo, está *implícito* en el tipo de herramienta formal usada para describir su dinámica (27) (ecuaciones diferenciales, operadores lógicos, autómatas finitos, lenguajes de programación). Si usamos matemática analítica, podemos hablar de excitación, inhibición, umbrales, sumas, productos por constantes, facilitaciones, derivadas (cambios espaciales y/o temporales en estas magnitudes), integrales (acumulaciones y «memorias» analógicas, procesos de carga y descarga de capacidades). Si introducimos elementos no lineales, tales como productos, exponenciales, zonas muertas, zonas cuadráticas o saturaciones, podremos hablar de oscilaciones no lineales, ciclos límite, etc... Análogamente, si introducimos variables lógicas podremos interpretar el disparo de una neurona, por ejemplo, como la afirmación de la «verdad» que le proponen sus dendritas. Si introducimos retardos, podremos hablar de periodos refractarios absolutos y relativos, de biestabilidad, de «estados lógicos» de «memoria» digital, etc...

Sin embargo, es posible pensar que el formalismo matemático necesario para modelar los aspectos más genuinos del comportamiento humano no está todavía disponible, como no lo tuvo la física hasta el cálculo diferencial de Newton. Esto nos lleva a la segunda limitación,

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

que hace referencia a los *significados* que no son computables, sino que se quedan siempre a nivel de conocimiento y en el dominio del observador externo. Los operadores *f* y *g* parten de variables formales (*x, m*) y producen nuevas variables formales (y), sin necesidad de referencia alguna al origen neurofisiológico del problema. Lo mismo nos da decir que estamos hablando de visión que de miedo; de percepción que de motoneuronas; de intención que de presión o temperatura. Todo ese conocimiento no computable (los campos semánticos de las entidades y relaciones del dominio) queda fuera del modelo por lo que hay que ser cuidadosos a la hora de interpretar resultados.

La tercera limitación hace referencia a los *errores* (voluntarios o no) asociados a *saltos de nivel*. En estos saltos del nivel fisiológico al nivel de conocimiento, por ejemplo, se asocian significados de un nivel de análisis alto (significaciones, propósitos, deseos, miedos o intenciones) a *variables formales* de un modelo que arranca a partir de una fenomenología de bajo nivel. Algunas posible «reglas de validez» del proceso de modelado computacional y subsiguiente *teorización* a partir de los resultados del modelo son (20):

R.1: *Selección* adecuada de la herramienta formal propia de cada nivel y especificación de las *suposiciones* implícitas a esta selección. Es decir, de las *limitaciones* que las hipótesis del modelo imponen sobre el campo semántico de interpretación de sus resultados. Es lícito usar ecuaciones diferenciales para modelar el potencial de membrana o para integrar la amplia variedad de registros a nivel de células ganglionares en retina (incluyendo en este caso retardos y operaciones no lineales locales en el espacio y en el tiempo). No es evidente en cambio el paso de las expresiones propuestas para relacionar trenes de impulsos a esquemas de percepción en vertebrados, por ejemplo. El lenguaje formal del modelo (los operadores) debe coincidir (estar próximo) con el de la naturaleza de los datos experimentales y con el de los experimentos propuestos.

R.2: *Coherencia interna,* acotando la anchura de la *banda semántica* en la que podemos movernos de forma legítima al modelar e interpretar resultados. Por ejemplo, no subiendo de forma injustificada la semántica. En muchas ocasiones se parte de una fenomenología compleja descrita a nivel de conocimiento (emoción, conciencia, intención, diagnóstico médico,...), se asignan variables y operadores del nivel lógico (verdad o falsedad de variables binarias que son transformadas por operadores del tipo «Y», «O», «NO») y se vuelve a interpretar los resultados en un nivel muy superior (cambios de motivación, diagnósticos clínicos «reales», ...). Se parte de descrip-

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

ciones bioquímicas en canales de sodio o potasio, potenciales de membrana, trenes de espigas, etc... y se proponen interpretaciones conductuales de carácter sintético global. El lenguaje formal de un modelo debe estar siempre en una banda semántica estrecha en torno a la naturaleza de los datos experimentales.

## 3. Aspectos Metodológicos: Niveles y Dominios de Descripción Comunes a Neurocienia y Computación

### *3.1. Niveles de Descripción de un Cálculo*

Decía Davis Marr que cualquier explicación de la percepción visual que se base sólo en el funcionamiento de las redes neuronales desde retina a corteza será absolutamente insuficiente. Lo que necesitamos tener es una «clara comprensión de lo que se debe calcular, cómo es preciso hacerlo, los supuestos físicos en los que se basa el método y algún tipo de análisis sobre los algoritmos que son necesarios para llevar a cabo ese cálculo» (11). Esto supone la introducción de un nivel adicional al que Marr llamó «Teoría del Cálculo» y Newell «Nivel de Conocimiento» (28).

Así, ahora es usualmente aceptado en el campo de la IA y en el modelado computacional en Neurociencia (3,7) que para analizar o sintetizar un modelo computable, tanto si el sistema es artificial como si es natural, es necesario distinguir e integrar al menos, tres niveles de descripción, tal como se ilustra en la figura 2:

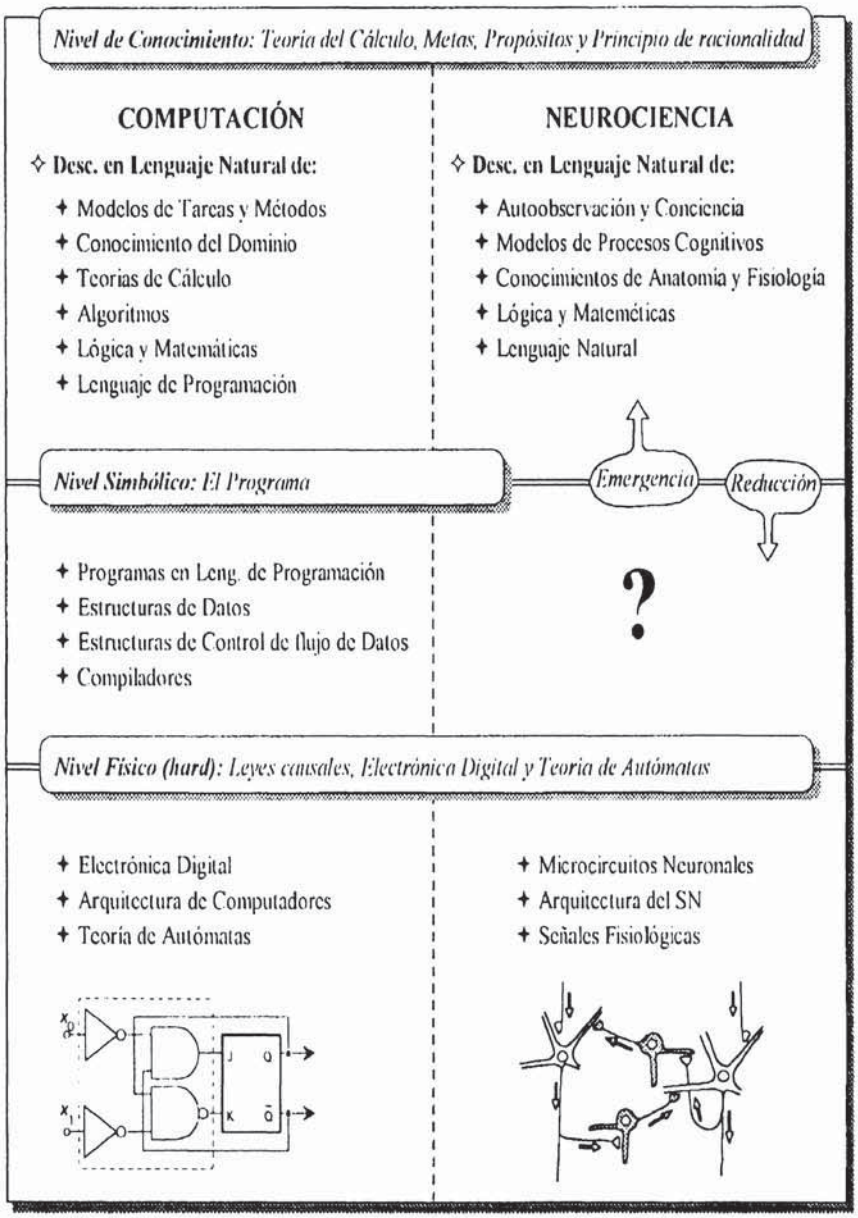
I) Una teoría de cálculo (Nivel de *Conocimiento* en Newell).
II) Un algoritmo (Nivel de los *Símbolos* —programa— en Newell).
III) Una implementación biológica o electrónica (Nivel *Físico* en Newell).

En el *primer nivel* tenemos los fundamentos teóricos del modelo que queremos hacer computable descritos en lenguaje natural y un posible esquema conceptual del proceso, junto con los conceptos propios del dominio. Para conseguir comprender un determinado proceso debemos empezar describiendo su propósito, las entidades y sus relaciones y algunas pistas sobre los principios en los que puede apoyarse. Este primer nivel de David Marr se engloba en el *nivel de conocimiento* propuesto por Newell en 1981. Nada podrá ser modelado computacionalemte si previamente no ha sido descrito de forma «clara, completa, precisa e inequívoca» en lenguaje natural y a nivel de conocimiento.

El *segundo nivel* de Marr (*representación* y *algoritmo*) incluye la descripción algorítmica del modelo anterior y se corresponde, aproxi-

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

FIGURA 2
Niveles de descripción de un cálculo igualmente válidos para la descripción del SN y la conducta emergente a partir del comportamiento de las redes neuronales. I. Nivel de Conocimiento. II. Nivel de los símbolos computables (lenguaje de programación) o neurofisiológicos. III. Nivel físico o fisiológico.




(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

madamente, con el *nivel de los símbolos* propuesto por Newell (el *programa*). La diferencia entre las propuestas de Marr y Newell es básica a nivel conceptual ya que lo que nos dice Newell es que prácticamente toda la parte relevante de una computación se queda en el nivel de conocimiento, incluyendo el algoritmo. Así, la propuesta de Newell es clara: «el nivel de la símbolos es el programa», todo lo demás queda «por encima» (el conocimiento) o «por debajo», (la máquina electrónica o la red neuronal).

El *tercer nivel* tiene que ver con todo el proceso de implementación que nos lleva del programa a los procesadores físicos o fisiológicos.

Visitemos ahora una supuesta «*máquina inteligente*» a la luz de la teoría de niveles, pensando en los razonamientos análogos que podría realizar un neurofisiólogo experimental. Encontramos primero un sistema físico, con una arquitectura determinada, donde navegan miles de señales binarias (0 ó 5 voltios) que son transformadas localmente por un conjunto de circuitos digitales (puertas lógicas y biestables, registros y contadores, multiplexos, etc...), de acuerdo con un esquema temporal marcado por un tren de impulsos. Esto es lo que nos contaría un ingeniero electrónico. Si la máquina que visitamos es una supuesta «*máquina emocional*», encontraríamos lo mismo. Así, el soporte físico de la supuesta emoción computable es el mismo que el del resto de los conceptos. Sólo encontramos circuitos electrónicos que evalúan, seleccionan, modifican y transfieren configuraciones binarias de dos niveles de tensión.

En cambio, si en la visita a la «*máquina inteligente*» incluimos a un *informático* cuya actividad más usual es la *programación*, sólo vería símbolos, variables e instrucciones de un lenguaje cuyo compilador ha producido la sucesión de configuraciones físicas que vio el ingeniero electrónico. Es decir, verá proposiciones matemáticas, operadores lógico-relacionales, variables de tipo lógico y valores permitidos para esas variables (verdadero o falso, «1» o «0»), y variables no numéricas tipo «ristras» (me gusta x = F, no me gusta = G) donde F y G, terminan finalmente en sus códigos ASCII o hexadecimal (F = 46, G = 47).

De nuevo nos parece evidente que en la visita de un informático a una «*máquina emocional*» no encuentra nada distinto de lo que encontró al visitar la «*máquina inteligente*» y, en ambos casos, nada diferente de lo que acepta un lenguaje de programación

Finalmente, si en la visita a esta «*máquina inteligente*» incluimos también un experto en «*ingeniería del conocimiento*», empezaría no estando especialmente interesado ni en la máquina física ni en el lenguaje de programación. Estaría preocupado por los *modelos* a nivel de conocimiento y por las *tablas de semántica* usadas para reescribir

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

las *entidades* y *relaciones* de esos *modelos* en términos de las primitivas del lenguaje de programación. ¿Dónde está pues la inteligencia?, ¿dónde está la emoción?, ¿dónde están los significados, los propósitos o las intenciones? Están en el *nivel de conocimiento* y en el *dominio del observador externo* (el *diseñador*) que *ha sido capaz de modelar* los aspectos más *relevantes de la inteligencia humana* hasta un nivel de detalle tal que las últimas entidades del último de los modelos ya pueden *identificarse* con los símbolos del programa. Sin embargo, el precio que ha tenido que pagar es dejarse toda la semántica fuera de la máquina y añadírsela al interpretar los resultados de la computación. Es decir, la máquina sigue siendo lo que siempre fue: un instrumento de cálculo que potencia y complementa las facultades del usuario, pero que depende de él para volver a tener semántica de lenguaje natural, de cognición y de emoción.

La *inteligencia* y la *emoción* se quedan en el *modelo* a nivel de conocimiento y en sus tablas de semántica que primero construye y después usa en la interpretación el observador humano. Este modelo y sus tablas de semántica son *independientes* del programa específico y de la máquina concreta en la que corre ese programa. Si el modelo es fino y profundo y ha sido capaz de captar los aspectos más genuinos del pensamiento humano y de la esfera emocional, podremos decir que el programa resultante mimetiza (simula) de forma razonable la parte computable de esa parcela del pensamiento humano.

Resaltemos ahora los siguientes puntos para resumir este apartado sobre el modelo de computación en un nivel y la conveniencia de usar tres niveles de descripción:

1. La existencia en cada nivel de un mismo esquema (espacio de entradas, espacio de salidas y reglas de transformación), con su semántica propia y con sus operadores característicos. Es decir, cerrado a *estructura* y *organización* con una causalidad propia del nivel. Al cambiar de nivel mantenemos la estructura del modelo pero cambiamos el significado de las entidades y sus relaciones. También cambiamos las leyes de casualidad, en general empobreciéndolas al bajar de nivel y enriqueciéndolas al subir. Cuando decimos que el potencial de membrana está relacionado con la corriente iónica y la conductancia de los canales por la ley de Ohm ($I_{Na^+} = g_{Na^+}(Vm - E_{Na^+})$), a nivel de conocimiento estamos diciendo todo lo que sabe un neurofisiólogo, pero a nivel de los símbolos sólo estamos diciendo que dos variables analógicas, (V e I) están relacionadas por un operador producto, sin referencia alguna a lo que puede ser una conductancia, y menos un canal iónico.

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

2. La inevitable reducción de conocimiento en los procesos de bajada de nivel y la correspondiente inyección en la interpretación de los resultados (emergencia). Este proceso reversible de *extracción/inyección* se realiza siempre a través de las tablas de semántica que establecen las correspondencias entre los espacios de entrada de niveles vecinos y entre los espacios de salida de esos mismos niveles.
3. Para entender el significado de la IA y del modelado computable en neurociencia es imprescindible «saber llevar la cuenta», distinguiendo de forma clara los significados asociados a cada nivel y no enriquecer artificialmente los resultados de un programa o de un modelo si no hay evidencia causal en los niveles inferiores. Es decir, al salir de un restaurante hay que ponerse el mismo abrigo que nos quitamos a la entrada, no entrar con harapos y salir con visones.

### *3.2. El Observador y los Dos Dominios de Descripción*

Queremos saber qué estamos diciendo de hecho cuando decimos que un modelo de retina es computable o cuando decimos que un programa de IA «diagnostica» como un cardiólogo o cuando decimos que un robot tiene propósitos y motivaciones en el medio. Para ello vamos a completar la descripción en tres niveles de todo modelo computable introduciendo la figura del *observador* externo a la computación (y al sistema biológico bajo análisis experimental) que superpone dos sistemas de referencia, dos dominios, sobre los tres niveles. Estos dos dominios de descripción son:

I) El dominio propio de cada nivel (*DP*)

II) El dominio del observador externo (*DOE*) que primero codifica y después interpreta el significado de la computación.

Tenemos así un edificio epistemológico de tres plantas (*conocimiento, símbolos* y *física*) y dos pisos por planta (*DOE* y *DP*) de forma que al modelar, programar e interpretar no debemos confundirnos ni de planta ni de piso. De lo contrario, los resultados del modelo o del programa de IA serán confusos y, probablemente, erróneos.

La introducción de la figura del observador y la distinción entre una fenomenología y su descripción, procede de la física y ha sido reintroducida y elaborada en el campo de la biología por Maturana (12) y Varela (36) y en la IA y la computación neuronal por Mira y Delgado (21,22,23).

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

En el dominio propio (*DP*), que se ilustra en la columna derecha del edificio de la figura 2, todo lo que ocurre en los distintos niveles es causal y no arbitrario. Las relaciones espacio-temporales entre los valores de las distintas variables son relaciones de *necesidad*. No pueden ser otras que las que su estructura determina. La semántica, además es propia e inherente al nivel. Estructura y función coinciden y ocurre «lo que tiene que ocurrir».

El *DP del nivel físico* es quizás el más evidente. En Electrónica, los inversores invierten y los contadores cuentan, porque el circuito está construido así. Las leyes son las de la lógica. Lo mismo ocurre en el *DP del nivel fisiológico*. El potencial de membrana, los trenes de espigas, los transmisores sinápticos, son entidades propias del nivel que se comportan «como tienen que comportarse», reaccionando ante perturbaciones de su medio con los cambios compensatorios que tienen impresos en su estructura. Las sinapsis, los contactos dendro-dendríticos y las neuronas hacen «lo que tienen que hacer» lo que las leyes anatomo-fisiológicas les exigen. El que un observador externo interprete aquello como percepción o dolor es sólo función de la posición del circuito en la red global, y pertenece al *DO*.

Si subimos ahora de los niveles físico y fisiológico al nivel de símbolos, vuelve a repetirse el proceso. El nivel de los símbolos en computación lo constituye el programa y ningún programa puede salirse de la sintaxis, la semántica y la pragmática del lenguaje de programación con el que ha sido escrito porque de lo contrario no sería aceptado por su compilador y, por consiguiente, no podría pasar al nivel físico. No podría ejecutarse. Lo que en el nivel físico eran niveles de tensión en circuitos digitales, ahora son valores de verdad (1 = verdadero, 0 = falso) en expresiones lógicas. Y esa es la única semántica del nivel.

Cuando hablamos de computación, el *DP* contiene a los niveles físico y simbólico. No creemos en la existencia de un nivel de conocimiento en el *DP* de un computador (su «mente») de la misma forma que no creemos en la existencia de una esfera emocional ni en la máquina ni en los programas. No pasa lo mismo cuando aplicamos la teoría de niveles y dominios a analizar las correspondencias entre el nivel fisiológico y el comportamiento. Aquí sí que hay nivel de conocimiento en el *DP* y, por consiguiente, cognición y emoción. Es decir, lo observado tiene las mismas características que el observador (lenguaje, pensamiento y emociones) y puede reflexionar sobre lo que piensa o siente.

Veamos ahora las descripciones correspondientes en el dominio del observador externo. Su existencia es evidente ya que estamos hablando

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

simplemente de la *persona que hace el modelo* y lo programa después. Las descripciones de una computación (o de lo que se observa en un experimento electrofisiológico) siempre se desarrollan en el *dominio del observador* (experimentador, analista, etc...), (*DO*), *en lenguaje natural y a nivel de conocimiento* usando la semántica propia del lenguaje natural y del conocimiento que tenemos sobre las entidades del dominio. Es decir, la primera versión en el proceso de síntesis de un modelo computable siempre se escribe en lenguaje natural, haciendo uso de las herramientas formales de las que disponemos (lógica y matemáticas) y con la semántica propia de lo humano. Cuando hablamos de presión, volumen o temperatura y escribimos PV=nRT, usamos una fórmula pero damos por supuesto, además, que P, V y T tienen el significado usual en Física. Si además estamos usando estas variables en un dominio médico damos por supuesto también el significado de «presión intracraneal», «volumen pulmonar» y «temperatura alta» (fiebre), porque esas variables tienen una nueva esfera semántica superpuesta a la que usaría un físico, por ejemplo.

Las descripciones de una computación en el *dominio del observador* (*DO*) siempre usan la *semántica* y la *causalidad* del lenguaje natural, incluyendo la lógica y las matemáticas y los metalenguajes del conocimiento específico de la tarea (i.e. diagnosis) y el dominio (i.e. cardiología) necesarios para especificar el modelo de conocimiento del problema que se quiere hacer computacional («un programa que diagnostique en cardiología»). Es evidente que en el *DO* también podemos hablar de las entidades y relaciones de los niveles físico y simbólico (hablaremos de inversores, álgebra de Boole, instrucciones de control, operadores lógico-relacionales, etc...) pero ahora no exigimos la causalidad del nivel de conocimiento, sino la propia de los símbolos y la electrónica digital, que es legítimo usarla para producir nuevas entidades del mismo nivel. El problema fundamental es no mezclar causalmente entidades que pertenecen al *DP* del nivel con otras entidades del *DO* que se usan para referenciar o explicar a las primeras.

En computación en general y en IA en particular la teoría de *niveles y dominios* de descripción de un modelo computable encuentra su uso natural en procesos de *ingeniería directa*, en los que partimos de un conjunto de especificaciones funcionales sobre la solución de un problema que queremos hacer computable, lo modelamos a nivel de conocimiento y construimos después un programa que, con el compilador adecuado, se ejecuta en una máquina física. Es decir, el mapa de niveles y dominios se recorre de izquierda a derecha y de arriba hacia abajo. Empezamos con un modelo en el *DO* y a nivel de conocimiento («3º izquierda») y

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

pasamos al *DP* a nivel simbólico («2° derecha»). Después el compilador la baja al *DP* del nivel físico («1° derecha»).

En neurociencia se recorre el camino a la inversa, haciendo el recorrido desde el nivel físico en el *DP*: Dada una red neuronal real, intentamos recuperar el «modelo» del que podían derivarse implementaciones como la que estamos analizando. Hacemos *neurofisiología inversa.*

## 4. De lo Natural a lo Artificial (de la Neurociencia a la Electrónica y la Computación)

La metáfora computacional ha sido usada insistentemente para intentar comprender las relaciones «*mente-cerebro*» aceptando de forma más o menos explícita que tales relaciones son isomorfas a las existentes entre «*programa y computador*». Ahora que hemos expuesto la perspectiva metodológica de niveles y dominios de descripción común a neurociencia y computación queremos creer que es evidente que esta analogía (mente-software, cerebro-hardware) no debe llevarse muy lejos por inapropiada. Por el contrario, creemos más adecuado el reconocer que ambos dominios (lo físico y lo biológico) pueden analizarse usando la misma metodología y las mismas herramientas formales, sin forzar la confusión de los campos semánticos propios de las entidades y las relaciones de ambos dominios (el de la computación y los cristales de silicio por un lado y el tejido biológico y la evolución por el otro).

Además, creemos que (sin necesidad de forzar la analogía) la relación puede ser fructífera en ambos sentidos. Es decir, buscando inspiración en «*lo natural*» para formular nuevos mecanismos y modelos de computación de utilidad técnica y usando lo que sabemos de «*lo artificial*» (Electrónica, Computación, Matemáticas y Teoría de Sistemas) para ayudar a comprender el funcionamiento del SN, la planificación de nuevos experimentos y la formulación de modelos que, al ser computables, nos van a permitir experimentar de forma nada cruenta.

Veamos ahora la primera parte de la conjetura:

*Es posible que la IA pueda obtener inspiración en lo natural para formular nuevos modelos computacionales de percepción, memoria, razonamiento, aprendizaje, planificación motora etc...*

Es decir, es posible que, usando cuidadosamente el esquema de niveles y dominios, podamos encontrar inspiración para el diseño de modelos útiles en IA aplicada a partir de las descripciones que

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

biólogos, bioquímicos, médicos, psicólogos, sociólogos y filósofos nos hacen de la fenomenología emergente del SN. Para ello, es tiempo de potenciar grupos de trabajo interdisciplinarios con físicos, matemáticos e ingenieros informáticos que complementen la perspectiva biológica para formular esos modelos computables. Este es el viejo sueño de los griegos de mecanizar los procesos del pensamiento (ahora decimos «hacer computables») que fue retomado en 1943 por Rosemblueth, Wiener y Bigelow (31) y McCulloch y Pitts (14), dando origen a la Cibernética y a la Biónica. El intento de «bio-inspiración» aparece de nuevo con el renacimiento del conexionismo en 1986 y está ahora en pleno auge con la búsqueda de «algoritmos genéticos». «arquitecturas evolutivas», «redes neuronales artificiales», «sistemas sensoriales neuromórficos», «robótica perceptual autónoma» y «arquitecturas intencionales».

Lejos de esta perspectiva, hay otras formas de considerar a la Biología como fuente inagotable de inspiración para la computación.

### *4.1. Nivel Físico*

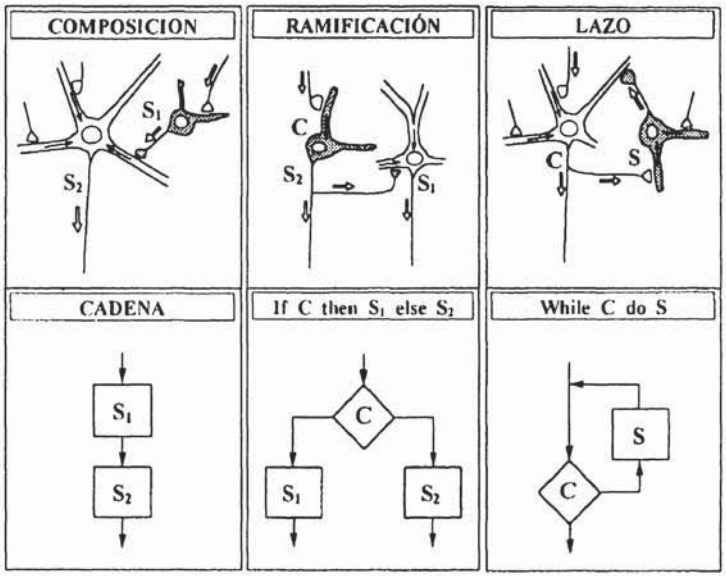
Muchas de las características de la anatomía y fisiología del SN, pueden usarse para el diseño de redes «neuronales» artificiales más potentes que las actuales (sumadores seguidos de sigmoides y «aprendizaje» por retropropagación del gradiente) (20). Tal es el caso de la computación logarítmica en sensores, las arquitecturas incrementales, la «poda» selectiva y los procesos cooperativos y otras formas de computación factorial que se inspiran en la alta tolerancia a fallos del SN donde, tras la lesión traumática o quirúrgica, el cálculo todavía permanece estable (1,10,6,5)

Si consideramos que la neurona es de hecho un gran integrador de información y aceptamos que el contacto sináptico tiene las características adecuadas para ser considerado un módulo básico (35,23), moviéndonos sólo por los campos dendrítico y axónico de una neurona ya encontramos una fuente inagotable de esquemas de conectividad (30,35). Por ejemplo, los tres necesarios para controlar el flujo de cualquier programa espacio-temporal: (1) *composición secuencial* (primero S1, después S2), (2) *ramificación condicional* (si se cumple C, entonces S1, si no S2) y (3) *lazos* (mientras se cumpla C, hacer S), tal como se ilustra en la figura 3.

Es posible que para ampliar las correspondencias entre funcionamiento de SN y computación tengamos que empezar a pensar en *computación bioquímica* y *mecano-cuántica*, por ser estos niveles mucho más finos que el eléctrico. De todas maneras, es en el problema de

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

FIGURA 3. Circuitos locales del bulbo olfatorio (27) asociados a las estructuras de control de micro-programas anatómicos.




la lesión física y la alta tolerancia que muestra al sistema nervioso para mantener estable la organización e, incluso, recuperar la función perdida mediante un proceso de rehabilitación (5) donde encontramos la mayor fuente de inspiración para la computación del futuro. La *idea general* es sencilla: No sólo podemos encontrar inspiración mirando a los circuitos de la máquina biológica, sino también observando su comportamiento global y reflexionando sobre *los principios de organización y estructura* que subyacen a esos comportamientos. Si somos capaces de formular estos principios, podremos diseñar sistemas de cálculo que los incorporen.

Los hechos concretos que nos sirven para introducir *el principio de organización cooperativa* y *tolerante a fallos* son también muy claros. Tras una lesión física del cerebro, resultado de una herida de guerra, un traumatismo o la cirugía, aparecen un conjunto de síntomas, acompañados por una disminución en la función global. Sin embargo, la función permanece (1,10,6). No hay nada parecido en las máquinas que nosotros construimos. No es frecuente que tras pegarle un tiro a nuestro computador personal y eliminarle un trozo del microprocesador y otro de la memoria RAM, siga funcionando aunque algo más lento y con algún error.

El *modelo* computacional que sugieren estos hechos es el de la computación factorial y los *procesos cooperativos* (19) donde se supone que cada neurona participa en muchos sistemas funcionales (SF) apor-

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

tando «factores» y que, a su vez, en cada sistema funcional participan factores procedentes de muchas redes. Al lesionar un área se elimina un factor pero la función permanece con los factores supervivientes.

### *4.2. El nivel de los Símbolos*

Aquí es donde creemos que la computación está más adelantada que nuestro conocimiento sobre la parte correspondiente del SN. Es decir, la parte más importante del edificio computacional está asociada al desarrollo de lenguajes de programación, compiladores, editores y algoritmos que nos permiten pasar de un conjunto de especificaciones funcionales (un modelo computable) a un programa. Sin embargo, existen pocos estudios sobre lo que podría ser un *lenguaje de símbolos neurofisiológicos*, capaz de trabajar sobre información poco estructurada y de orígenes múltiples. Sigue siendo un desafío el comprender y modelar la habilidad del SN para organizar la información presentada de forma masiva y poco estructurada, produciendo jerarquías, orden, asociación, recuperación dinámica selectiva y todo el resto de los «algoritmos y programas» sobre los que se basa el pensamiento, la emoción, el lenguaje y la acción. Esta habilidad es la que hace que los seres vivos no necesiten programación, sino aprendizaje. Por el contrario, la falta de esta habilidad es la que hace que los computadores necesiten que todo les sea programado, incluidos los llamados algoritmos de «aprendizaje» (33).

### *4.3. El nivel del Conocimiento*

La fenomenología de este nivel incluye todo lo que la psicología, la lingüística, la sociología, la filosofía y el resto de las «humanidades» han aportado al estudio de la mente. De aquí, la IA se ha aprovechado hasta ahora sólo de algunas de las estrategias de «*solución de problemas*» que, a su vez, son una parte mínima de los procesos cognitivos. Todo modelo computable necesita primero una formulación en lenguaje natural clara, completa, precisa e inequívoca. Por consiguiente, las soluciones computables han existido siempre previamente como soluciones humanas, «*con papel y lápiz*». Queda fuera del alcance de este trabajo el revisar con mayor extensión y profundidad esta vía de conexión.

## 5. De lo artificial a lo natural

Veamos ahora la otra conjetura complementaria (¿cómo pueden ayudar la Física, la Teoría de Sistemas y la Computación a la comprensión del SN?):

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

*El estudio experimental del SN y la teorización sobre las relaciones mente-cerebro pueden sacar inspiración y aprovecharse de las técnicas, los métodos y los resultados usuales en la Computación (incluida la IA), la Electrónica y la teoría de Sistemas.*

### *5.1. El Nivel Físico*

El análisis de lo natural puede mejorarse substancialmente si se reflexiona sobre los métodos de la Física y la Computación usando la metodología de niveles y dominios descrita previamente, sólo que ahora debemos recorrerla de abajo hacia arriba, haciendo Neurofisiología inversa (16,17). Es decir, partiendo de una red neuronal e intentando recuperar el modelo. Este problema general de análisis se puede plantear en los siguientes términos.

«Dado un conjunto de circuitos y señales de los que conocemos parcialmente sus relaciones causales en el nivel biofísico y/o bioquímico, encontrar:

1º *Una formulación de las interdependencias causales de esas señales en el DP del nivel físico, usando las herramientas formales de la Teoría de Sistemas y las Matemáticas. Es decir, formular modelos computables sin salir del nivel físico.*

2º *Un conjunto de símbolos neurofisiológicos (los «roles» que desempeñan las señales) que intervienen en la descripción de esos procesos al nivel de los símbolos y un conjunto de «algoritmos» que explique esas relaciones.*

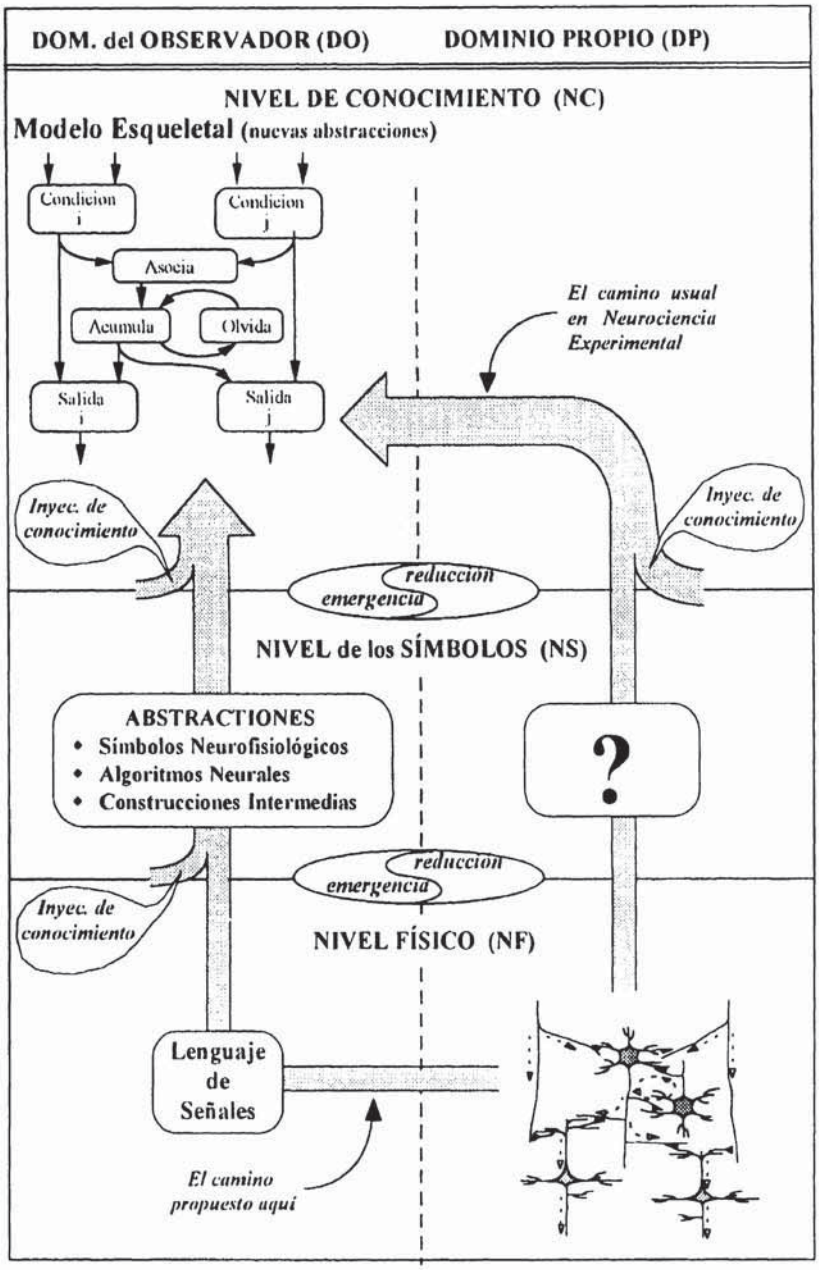
3º *Un modelos a nivel de conocimiento y en el DO, a partir de las cuales el buen Dios y su aliada la evolución, podrían haber diseñado una red neuronal funcionalmente análoga a la que estamos analizando.*

En la figura 4 presentamos la trayectoria de la Neurofisiología inversa sobre el esquema de niveles y dominios. Partimos del «bajo-derecha» (*DP* del nivel físico) y queremos llegar al «segundo-izquierda» (nivel de conocimiento en el *DO*). Hemos usado un circuito de reflejos condicionados para ilustrar la descripción. Su uso por Ledoux en el estudio del miedo y su relación con la emoción ha influido en nuestra selección (2,8,9). En otras ocasiones hemos usado circuitos de la retina (17) o esquemas de inhibición lateral recurrente (23).

La estructura de cálculo asociada al comportamiento reflejo ilustra el problema del análisis porque se reproduce a nivel de comportamiento. Es decir, no es difícil establecer la analogía entre la descripción del cambio de conducta que experimenta un animal durante el proceso de condicionamiento y la descripción equivalente del funcionamiento

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

FIGURA 4. Ilustración del uso en Neurociencia experimental del marco conceptual de «niveles y dominios» de descripción para el caso del arco reflejo.




(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

de un circuito neuronal que puede soportar ese comportamiento. El arco reflejo realiza «computación acumulativa», permitiendo asociar configuraciones sensoriales neutras con otras relevantes para la supervivencia, acumulando las asociaciones espacio-temporales (lo próximo, lo análogo), abriendo vías temporales de asociación con patrones de respuesta y ampliando o extinguiendo estas vías resultado de la integración temporal entre percepción y acción.

El análisis comienza identificando el soporte anatomofisiológico con las técnicas usuales (bioquímicas, farmacológicas y moleculares; métodos morfológicos y fisiológicos de registros intra y extracelular, métodos combinados etc...) (32). La perspectiva de niveles y dominios nos aconseja distinguir claramente entre lo medido y lo interpretado, es decir separando las señales y procesos del *DP* del conocimiento adicional usado en la formulación del modelo. Nos encontramos con un repertorio de señales y estructuras locales de módulos y esquemas de conectividad. Para el arco reflejo encontramos circuitos del tipo del de la parte inferior derecha de la figura 4. La presencia de señal en la vía $E_i$ activa un patrón de respuesta $R_i$. Si coinciden los estados de actividad de la $E_i$ con los de la otra línea $E_j$ (en principio neutra en relación al patrón de respuesta $R_j$) y se acumula esta persistencia hasta un cierto valor umbral, la señal $E_j$ también producirá $R_i$.

Hay un proceso de extinción siempre activo que desconecta funcionalmente $R_i$ de $E_j$ si no persiste la asociación $(E_i, E_j)$.

Así, terminado el análisis del nivel físico parece evidente que necesitamos:

*Un lenguaje de señales neurofisiológicas,* (electrónicas, bioquímicas, biofísicas), con herramientas formales adecuadas para la descripción de los potenciales de membrana, las espigas y los procesos de excitación-inhibición.

En el nivel físico, donde las señales son potenciales lentos o trenes de espigas, las coincidencias se modelan con *productos* y la persistencia con *integrales*, y podríamos escribir el siguiente modelo analógico:

$$R_i(t) = A_{ii} \cdot E_i(t) + \underbrace{K^i_2 \cdot \left( \int_{t-T}^{t} E_i(\tau) \cdot E_j(\tau)\, d\tau \right)}_{A_{ij}} \cdot E_j(t)$$

$$R_j(t) = A_{jj} \cdot E_j(t) + \underbrace{K^j_2 \cdot \left( \int_{t-T}^{t} E_j(\tau) \cdot E_i(\tau)\, d\tau \right)}_{A_{ji}} \cdot E_i(t)$$

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

Esta formulación analógica de los reflejos condicionados muestra como es la historia reciente de estímulos (coincidencias $E_i(\tau)$, $E_j(\tau)$, dentro del intervalo $T$), la que abre o cierra las vías de asociación entre estímulos y respuestas. Los coeficientes $A_{ij}$ y $A_{ji}$ son los responsables del aprendizaje, y se asocian a la «eficacia sináptica», pero nada sabemos de la semántica de los estímulos ($R_i$, $R_j$), salvo su carácter ordenado (a nivel de conocimiento) en una escala (dolor-huída, neutralidad, placer-atracción). Así termina la descripción del nivel físico en el *DP*. Si llamamos miedo, campanilla, comida, estímulo neutro o estímulo aversivo a las variables ($E_i$, $E_j$, $R_i$, $R_j$) ya estamos interpretando el experimento en el *DO* y a nivel de conocimiento.

Usualmente los resultados del análisis en el nivel físico se formulan en términos de modelos puntuales «sin morfología» (15,34), o más detallados a nivel biofísico (potencial de membrana, red de generadores dependientes y conductancias en paralelo, etc...), hasta llegar al detalle dendro-dendrítico (35). Sin embargo dos reflexiones suelen olvidarse a la hora de determinar el valor del modelo:

1. La *capacidad explicativa* del modelo sobre SN está siempre limitada «a priori» por la naturaleza de las *herramientas* formales usadas para formularlo (ecuaciones diferenciales y lógica).
2. Para comprender el significado de una señal fisiológica, es necesario conocer su *historia*, el contexto del modelo. Es decir, el conjunto de recodificaciones que ha sufrido desde la vía sensorial o por retroalimentación. Las mismas señales pueden corresponder a símbolos diferentes.

### *5.2. El Nivel de los Símbolos*

Para subir el modelo de nivel y convertirlo en un modelo estructural a nivel de conocimiento tenemos que abstraer y generalizar, buscando los símbolos resultado de la abstracción de la variable física que lo soporta. Se puede pasar así de operaciones analíticas (tales como sumas o productos) a verbos de los que ese modelo analítico es sólo un caso particular, obteniendo así esquemas como el de la parte superior izquierda de la figura 4, que es una descripción en lenguaje natural del proceso:

*«calcular condición-i», «calcular condición-j», «asociar espacio-temporalmente, i-j», «acumular-olvidar», «disparar patrón de respuesta-i», «disparar patrón de respuesta-j».*

Donde *«calcular condición-i»* es una inferencia que representa todo el cálculo previo necesario para obtener el estado de actividad de la línea i. Este preproceso,

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

puede incluir desde la simple actividad directa de una modalidad sensorial hasta un complejo proceso de reconocimiento de caracteres.

El siguiente verbo del modelo es «*asociar i-j*» y tiene que ver con cualquier procedimiento que permita medir la vecindad espacio-temporal de dos símbolos neurofisiológicos. Es posible detectar distintos mecanismos de asociación (sumas, productos, modulaciones, coincidencias lógicas, etc...) responsables en el nivel físico de este proceso. Y lo mismo podríamos analizar el significado computacional del resto de los verbos («*acumular*», «*olvidar*», «*disparar patrón Ri, Rj*»), mediante procesos sinápticos o redes de interneuronas.

Si seguimos el camino ascendente en la figura 4 (en el sentido de la Neurofisiología inversa), ahora tendríamos que formular el *programa*. Es decir, la descripción de la *computación* en términos de *símbolos neurofisiológicos*. Para ello nos hace falta en el *Nivel de los Símbolos*:

> Un conjunto de abstracciones desde el nivel de las señales fisiológicas, hasta el nivel de los símbolos neurofisiológicos. Estas abstracciones deben ser independientes de las implementaciones anatómicas concretas y de las señales que las codifican.

En computación no se pasa directamente de la electrónica digital al lenguaje natural sino a través del nivel intermedio de los símbolos usados por los lenguajes de programación. Sin embargo, no existen propuestas equivalentes en Neurociencia para este nivel intermedio. Hay más datos que teorías integradoras. El simbolismo, en neurología y en computación, siempre nace en el dominio del observador externo. En el dominio propio sólo hay señales y tejido. Los *Símbolos en el DP:*

*Son:* Configuraciones específicas de señales espacio-temporales (eléctricas, químicas y electrónicas), («*llaves*»), con un referente en el medio externo o interno del organismo, y las correspondientes estructuras anatomofisiológicas («*puertas*» abiertas por esas «*llaves*»).

*Actúan:* Estas «llaves» neuronales actuan como enlaces dinámicos y han sido adquiridas (anatomica y funcionalmente programadas) por la evolución y la genética o por el aprendizaje. Representan (sustituyen) al referente externo en todos los procesos de información subsiguientes.

En cambio los *símbolos en el DO.*

*Designan:*

a) Entidades del medio relevantes para la supervivencia.
b) Relaciones multimodales y temporales entre estas entidades.
c) Conceptualizaciones primarias (señales de alerta, homeóstasis, ...)
d) Reacciones compensatorias.

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

s) Estabilidad de la especie (símbolos sexuales, de agresión o escape, ..., descriptores de necesidades internas, sueño, sed, ...).

*5.3. Nivel de Conocimiento*

Finalmente, para recuperar el modelo nos hace falta en el *Nivel de Conocimiento:*

Un nuevo conjunto de *abstracciones*, desde el nivel de los *símbolos neurofisiológicos* hasta el nivel de conocimiento cuya ontología da lugar a las descripciones en lenguaje natural de lo que llamamos «*actividad emocional*» o «*comportamiento inteligente*». De nuevo, estas segundas abstracciones deben ser independientes del simbolismo y, a su vez, independientes del nivel físico.

Supongamos que damos por finalizado el análisis del nivel físico, que ya sabemos todo lo referente a las *señales* y los *operadores* que las transforman. Es decir, que disponemos de una teoría neuronal completa a nivel físico, de forma análoga a cómo los físicos y los ingenieros electrónicos conocen la Electrónica Digital y la Arquitectura de Ordenadores. ¿Conoceríamos ya lo que está calculando la máquina?, ¿conoceríamos los procesos cognitivos emergentes de las redes neuronales?, ¿conoceríamos los procesos emocionales?. Claramente, no. Del sólo conocimiento del nivel físico no se puede obtener la descripción de la computación en los otros niveles, porque un mismo modelo puede reducirse usando distintos algoritmos y programas y un mismo programa puede ejecutarse en máquinas diferentes. Así, la correspondencia no es biunívoca y al igual que se pierde conocimiento en la reducción hay que inyectarlo en la interpretación. Es decir, cualquier explicación de los procesos mentales no puede basarse sólo en el funcionamiento de las redes neuronales, sino que necesita ser complementado, al menos por una clara comprensión de los símbolos y las entidades del nivel de conocimiento, en términos de cultura, historia, civilización y evolución en el medio.

## 6. La Emoción en la IA: ¿Existe la Máquina Emocional?

Ahora que hemos visto los aspectos metodológicos usados por la computación y la neurociencia para comprender el significado de un programa o los resultados de un experimento, distinguiendo claramente las entidades que hacen referencia al dominio propio de las que introduce el observador externo y distinguiendo también entre los tres niveles de descripción de una misma fenomenología, creo que es el momento

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

de reflexionar sobre los conceptos de la esfera emocional y su relación con la IA.

Si seguimos el camino análogo al seguido en la esfera cognitiva deberíamos hablar ahora de *emoción artificial,* sentimientos artificiales, dolor y amor artificiales etc. Es decir, deberíamos hablar de modelos y programas que emulan los aspectos más genuinos de la conducta emocional. Ya anunciamos al comienzo del trabajo que esta perspectiva no nos parece útil ni necesaria a una parte importante de los profesionales de la IA que, muy al contrario, creemos que nos distrae a Neurólogos e Informáticos en el camino duro de encontrar beneficios mutuos en nuestros intercambios de métodos, técnicas y conceptos.

En el camino desde la Electrónica y la Computación a la experimentación y teorización en el campo de la emoción, sí que creemos que pueden ser muy útiles las técnicas de modelado computacional y la simulación mediante programas para ayudar a comprender los resultados experimentales en el estudio de la emoción. Es decir, la IA es una buena técnica de análisis de la emoción natural. De hecho, tal como comentamos en la sección 2.2, los *significados no son computables*, sino que se quedan siempre a nivel de conocimiento y en el dominio del observador externo a la computación. Así pues, los operadores $f$ y $g$ enlazan los espacios de representación de las variables formales $(x,m)$ con las $(y)$ y lo mismo podemos decir que enlazan emociones como que enlazan potenciales en espiga o niveles de iluminación en fotorreceptores.

Conviene recordar aquí el *error* frecuente que también mencionamos en el apartado 2.2., de los *saltos de nivel* (desde el físico o simbólico al de conocimiento) y *los saltos de dominio* de descripción (desde el propio al del observador y viceversa), asociando entidades que pertenecen a dominios cognitivos distintos. Gran parte de las controversias relacionadas con el tema *mente-cerebro* (25) también se aclararían usando la teoría de niveles y dominios para descubrir las cuestiones subyacentes a preguntas planteadas de una forma poco adecuada.

La otra parte de nuestra conjetura (la IA puede aprovecharse del estudio de la emoción, de la misma forma que ha sacado provecho del estudio de los procesos cognitivos) no la veo útil. Creo que la mayor parte de las importaciones que se han realizado desde la esfera emocional hasta el campo de los computadores y la IA hablando de «*emoción artificial*» o de «*máquinas que sienten*» no tienen fundamento y, además, se trata sólo de etiquetas que se añaden de forma gratuita al formular e interpretar las funcionalidades de un programa pero que no tienen contrapartida real en el dominio propio de la computación. Dicho de otra forma, el

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

modelo computacional que subyace a términos tales como motivación, propósito o intención es siempre un conjunto de tablas, un autómata finito, un grafo y un conjunto de variables lógicas. Mucho más arbitrario es el uso de términos tales como conciencia, alegría, miedo, ansiedad, deseo , etc... que sólo caben en el campo de la novela de ficción.

En el análisis paralelo que hemos realizado en el apartado 3 de la «*máquina inteligente*» y «*la máquina que siente*» creemos que ha quedado claro el error de interpretación de estos conceptos para los que se usa el mismo símbolo (la misma palabra) cuando hacen referencia a humanos (símbolo «caliente», con semántica del lenguaje natural y de la neurofisiología por ejemplo «*propósito*») que cuando hacen referencia a la codificación del símbolo (símbolo «frío», por ejemplo «*propósito*»), que en este segundo caso se trata de un autómata finito y un conjunto de expresiones lógicas que hay que evaluar para decidir sobre los cambios de estado de ese autómata. Obsérvese que en este segundo caso la causalidad es también totalmente distinta. La conclusión aparente de este apartado es que no hay evidencia de que exista nada residente en una máquina o en un programa que nos permita hablar de máquinas emocionales.

## 7. Los Propósitos en IA: Existe la Máquina Intencional?

¿Existen propósitos artificiales? ¿Dónde reside la intencionalidad de un robot? Preguntas de esta naturaleza son frecuentes en IA. De hecho, si tuviéramos que elegir un concepto representativo de la frontera entre mente y computación este sería el concepto de intención y su vecino semántico el concepto de propósito. Los aspectos más genuinos del comportamiento humano, liderados por el lenguaje, están caracterizados por su naturaleza intencional. Por eso, desde la introducción del «LEKTON» por los estoicos la formulación de un cálculo intencional, que pudiera operar sin hacer referencia a la descripción en extenso de las clases ha sido buscado insistentemente. Se ha dado por supuesto que los propósitos y las intenciones, al igual que los significados de los conceptos, pueden ser *propiedades residentes* en un sistema físico artificial como una computadora. En nuestra opinión no hay evidencia de este hecho y queremos creer que al lector de este trabajo, a la luz del esquema de niveles y dominios, no le resultará difícil aceptar que estos conceptos de propósito e intención son de hecho:

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

*Construcciones teóricas creadas sobre el lenguaje natural por un observador externo y a nivel de conocimiento (o por autoobservación consciente) para facilitar la descripción del comportamiento observable en la interacción adaptativa de un sistema con su medio.*

La contrapartida física y a nivel de los símbolos en el *DP* de estos conceptos sigue las reglas generales que hemos descrito previamente. Es decir, no hay nada específico en la intencionalidad que pueda permitirnos identificarla como algo residente en el hardware y/o en el software de un sistema, fuera de los esquemas organizacionales y estructurales comunes al resto de los conceptos modelados. Cuando hacemos una descripción intencional de un sistema y después queremos encontrar una versión computable del mismo toda la semántica se queda en el *DO* y a nivel de conocimiento.

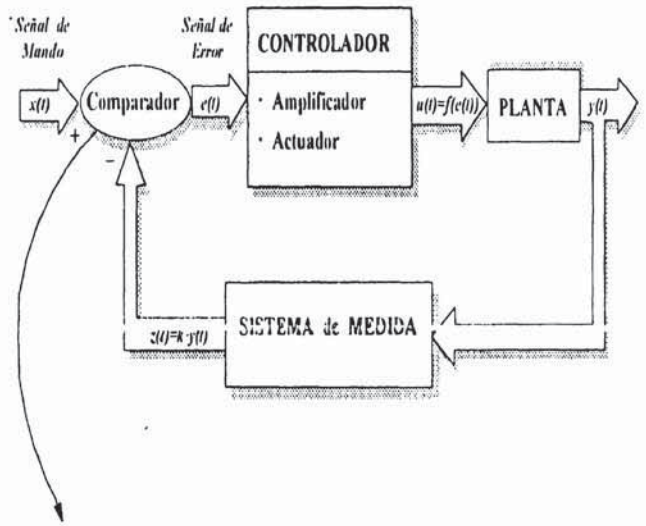
En los últimos 50 años la IA ha luchado por modelar versiones computacionales de algunos de los términos que usamos los humanos para describir métodos para resolver problemas técnicos («buscar», «comparar», «ordenar», «seleccionar», «calcular», «establecer», «refinar», «clasificar», etc...) pero los verbos cognitivos y emocionales más característicos de lo humano («pensar», «imaginar», «sentir», «creer», «intentar», «esperar», «desear», «temer», «odiar», «amar», ...) todavía no tienen modelos computables. Decir que una máquina tiene intenciones y propósitos en un medio es sólo una forma de hablar que queda vacía de contenido al llegar a la frontera de un compilador. Vamos a analizar el concepto de propósito por ser el más limitado de todos los anteriores y previo a abordar el problema de la intencionalidad.

Si un observador afirma que un robot tiene *propósitos* en el medio, lo que nos está diciendo es que en el hardware y en el software de ese robot existe una *estructura de control* basada en uno o varios lazos de realimentación negativa como el que se ilustra en la figura 5, en los que una parte del sistema realiza funciones «sensoriales» que monitorizan la evolución temporal de un conjunto de magnitudes físicas (su «percepción» del medio) y de otro conjunto de efectores ( motores que desplazan ruedas, etc) que el propio sistema tiene en ese medio y que participan en la acción que está realizando el robot (navegar sin tropezar con un obstáculo, coger un objeto, etc.). En ambos casos estamos hablando de series temporales de valores numéricos en un conjunto de magnitudes físicas.

Existe en la máquina otro hardware y/o software encargado de comparar (restar) el resultado del «report» de estos sensores para medir la discrepancia (el «error») entre la representación sensorial de la respuesta real y el estado interno de un circuito que representa la respuesta

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

FIGURA 5. Diagrama de bloques de un sistema de control usado para ilustrar la idea de *propósito* cuando se usa en el *DP* para hacer computable el concepto de *propósito* a nivel de conocimiento y en el *DOE*.




| Sub. Tarea | Método | Representación del conocimiento | Inferencia |
|---|---|---|---|
| **Comparación** | • Algoritmo de Resta | • Vectores Lógicos | • Deductiva (Reglas) |

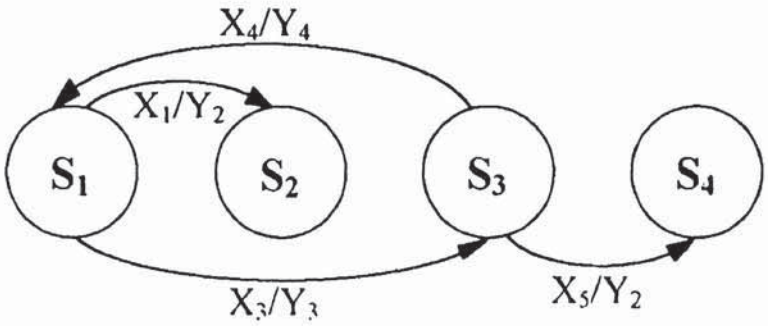
deseada, la meta (un valor concreto) o el propósito (un conjunto de valores y una secuencia de control del tipo «*si la variable x cumple la condición y, entonces emprender acción z, si no ....*»). Obsérvese de nuevo que *meta* y *propósito* son términos lingüísticos del *DO* que cuando pasan a nivel de los símbolos del *DP* se convierten en conjunto de variables y expresiones que, con el mismo nombre (*meta y propósito*) sólo son secuencias de valores numéricos que miden estados de actividad en circuitos electrónicos.

Sigamos con la descripción de la supuesta «*máquina intencional*» que maneja propósitos. Como consecuencia de la discrepancia en el proceso de comparación (resultado de la evaluación de un conjunto de expresiones aritmético-lógicas) se actúa sobre los efectores para corregirla. Así, cuando decimos que un robot tiene *propósitos* lo que estamos diciendo es que es capaz de *medir, realimentar, comparar,* y *actuar* sobre los efectores de acuerdo con un esquema de realimentación negativa como el que se ilustra en la figura 5.

Cuando un observador externo ve que al perturbar la conducta del robot este tiende a cancelar esa perturbación volviendo a la trayectoria inicial e intentando llegar a unas determinadas coordenadas

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

decimos que su conducta está «*guiada por propósitos*» pero detrás de esos supuestos propósitos sólo hay un entramado de *lazos de realimentación múltiple* y *una secuencia de instrucciones de control* que describen, a nivel de los símbolos, un plan como el que se ilustra a continuación.




*«Si estando en $S_1$ aparece una situación caracterizada por $X_1$, pasar a $S_2$ y ejecutar $Y_2$. Si aparece $X_3$, pasar a $S_3$ y ejecutar $Y_3$. Si estando en $S_3$ aparece $X_4$, volver a $S_1$ y ejecutar $Y_4$. Si estando en $S_3$ aparece $X_5$ pasar a $S_4$ y ejecutar $Y_2$»*

Es decir, detrás de un propósito hay un grafo o un autómata finito que describe, en extenso, las transiciones y acciones a llevar a cabo ante cada configuración de entrada. De forma análoga, detrás de una intención hay un conjunto de grafos. En ambos casos estamos hablando de algoritmos y heurísticas de control de los que toda la navegación instrumental de la moderna aviación está repleta y nadie suele afirmar que el avión x tiene la intención o el propósito de aterrizar en el aeropuerto y. Sus estados meta son configuraciones particulares en las variables físicas (presiones, alturas, coordenadas distancias, niveles de combustible, etc...) que caracterizan su estado y el estado del entorno. Somos nosotros, al modelar el problema, los que introducimos el concepto de intencionalidad en el sentido de *conocimiento estratégico* que supervisa la combinación de métodos para la solución de un problema.

## 8. Reflexión final

En este trabajo hemos comentado los términos en los que creemos que la neurociencia experimental puede sacar provecho del mundo de la electrónica y la computación lejos de los triunfalismos usuales en

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

la inteligencia artificial, por una parte, y de la visión a veces limitada de la experimentación en neurociencia que termina siempre en las fronteras del nivel físico o salta bruscamente y no siempre con fundamento desde el nivel físico al lenguaje natural. La consideración del nivel de conocimiento y el uso de los procedimientos de análisis usuales en física y computación, con la distinción entre las entidades del dominio propio y las del dominio del observador, ayudan a aproximar la neurociencia experimental al campo de la computación en general y de la IA en particular. Es prudente mantenerse lejos de las visiones optimistas que afirman que pensamiento y computación coinciden (29) y de las puramente estructurales y compensatorias (12,36), haciendo uso de los modelos computacionales mientras nos sean útiles, hasta que el avance en la evidencia experimental los haga innecesarios.

En cuanto al uso de términos de la esfera emocional mi posición se caracteriza por el rechazo a usar excesivos conceptos antropomorfos, sacados de la biología y las ciencias del comportamiento, para designar entidades y relaciones del mundo de lo artificial de semejanza sólo aparente a lo designado por esos mismos conceptos cuando se refieren al dominio de lo vivo. Es decir, no creo adecuado hablar de *robots intencionales,* por poner un ejemplo, porque creo que la riqueza semántica del concepto de intención en humanos se pierde en gran parte (si no toda) al pasarla a un modelo computable, donde termina en una tabla o en *un autómata finito* cuyas variables son *expresiones lógicas* de las que, a nivel de los símbolos, sólo se puede afirmar su verdad o falsedad en extenso.

## Agradecimiento

Agradezco el soporte de la CICYT a través del proyecto TIC - 97-0604 en cuyo contexto se han realizado parte de los trabajos que aquí se mencionan.

## Bibliografía

1. Beach, F.A. et al. (1960): The Neuropsychology of Lashley. McGraw-Hill.
2. Beeman, M.; Ontary, A.; Monti, L.A. (1995): Emotion-Cognition Interactions. In Arbib, M.A. (ed.): Handbook of Brain Theory and Neural Networks. The MIT Press. pp.360-363.
3. Churchland, P.S.; Sejnowski, T.J. (1992): The Computational Brain. MIT Press.
4. Craik, K. (1943): The Nature of Explanation. Cambridge University Press. Cambridge.

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

5. Delgado, A.E. (1978): Modelos Neurocibernéticos de Dinámica Cerebral. Tesis Doctoral. ETSIT. Madrid.
6. Gonzalo, J. (1952): Las Funciones Cerebrales Humanas según Nuevos Datos y Bases Fisiológicas. Instituto Cajal de Investigaciones Biológicas, Vol. XLIV, Madrid.
7. Hawkins, H.; McMullen, T.A. (1996): Auditory Computation: An Overview. In Hawkins, H. et al (eds.): Auditory Computation. Springer. pp. 1-14.
8. LeDoux, J.E. (1995): In Search of an Emotional System in the Brain: Leaping from fear to Emotion & Consciousness. In Gazzaniga, M.S. (ed.): The Cognitive neurosciences. The MIT Press. pp. 1049-1061.
9. LeDoux, J.E.; Fellous, J.M. (1995): Emotion and Computational Neuroscience. In Arbib, M.A. (ed.): Handbook of Brain Theory and Neural Networks. The MIT Press. pp.356-359.
10. Luria, A.R. (1974): El Cerebro en Acción. Ed. Fontanella, Barcelona.
11. Marr, D. (1982): Vision. Freeman, New York.
12. Maturana, H.R. (1975): The Organization of the Living: A theory of the Living Organization. Int. J. Man-Machine Studies, 7, pp.313-332.
13. McCulloch, W.S. (1965): Embodiments of Mind. The MIT Press. Cambridge, Mass.
14. McCulloch, W.S., and Pitts, W. (1943): A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, 5 pp. 115-133.
15. McCulloch, W.S.; Arbib, M.A.; Covan, J.D. (1962): Neurological Models and Integrated Processes. In Yovits, M.C. and Cameron, S. (eds.): Self-Organizing Systems. Spantan Books, pp. 49-59.
16. Mira, J. (1996): «Reverse Neurophysiology: The "Embodiments of Mind» Revisited. In Moreno-Díaz R.; Mira-Mira J. (eds.): Brain Processes, Theories and Models. The MIT Press, Mass. pp. 37-49.
17. Mira, J. (1998): Operaciones «Inteligentes» en Sistemas Artificiales: La Perspectiva de la Inteligencia Artificial en la Comprensión del Sistema Nervioso. Revista de la Real Academia de Medicina de Catalunya. Vol 12. Sup. 1. pp. 87-107.
18. Mira, J. et al (1995): Aspectos Básicos de la Inteligencia Artificial. Sanz y Torres. Madrid. pp. 485-575.
19. Mira, J. et al. (1995): Cooperative Processes at Symbolic Level in Cerebral Dynamics: Reliability and Fault Tolerance. In Moreno R. and Mira, J. (eds.): Brain Processes, Theories and Models. The MIT Press. pp. 244-255.
20. Mira, J., Delgado, A.E., Alvarez, J.R., de Madrid, A.P., and Santos, M., (1993): Towards More Realistic Self-contained Models of Neurons: High-order Recurrence and Local Learning. In: Mira, J., Cabestany, J., and Prieto, A. (eds.): New Trends in Neural Computation. LNCS-686, Springer-Verlag, , pp. 55-62.
21. Mira, J.; Delgado, A.E. (1987); Some Comments on the Antropocentric Viewpoint in the Neurocybernetic Methodology. Proc. of the Seventh International Congress of Cybernetics and Systems, Vol. 2. London. pp. 891-895.
22. Mira, J.; Delgado, A.E. (1995): Computación Neuronal Avanzada: Fundamentos Biológicos y Aspectos Metodológicos. En Barro, S.; Mira, J. (eds.): Computación Neuronal. Cap. VI, Servicio de Pub. e Intercambio Científico. Univ. de Santiago de Comp.. pp. 125-178.
23. Mira, J.; Delgado, A.E. (1997): Some Reflections on the Relationships Between Neuroscience and Computation. In Mira, J.; Moreno-Díaz, R.; Cabestany, J. (eds.):

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es

Biological and Artificial Computation: From Neuroscience to Technology. LNCS, 1240. Springer-Verlag, Berlin. pp. 15-26.
24. Mira, J.; Moreno-Díaz, R. (1984): Un Marco Teorico para Interpretar la Funcion Neuronal a Altos Niveles. En Moreno-Díaz, R.; Mira, J. (eds.): «Biocibernetica: Implicaciones en Biologia, Medicina y Tecnologia». Siglo XXI Editores, SA. Madrid. pp. 149-171.
25. Mora, F. (1995): «Neurociencia y el Problema Cerebro-Mente». En F. Mora (ed.): El Problema Mente-Cerebro. Alianza Universidad. pp. 261-288.
26. Moreno-Díaz, R. (1997): Systems Models of Retinal Cells: A Classical Example. In J. Mira et al (eds.): Biological and Artificial Computation: From Neuroscience to Technology LNCS, 1240. Springer. Berlin. pp. 178-194.
27. Moreno-Díaz, R.; Mira, J. (1996): «Logic and Neural Nets: Variations on Themes by W.S. McCulloch». In Moreno-Díaz, R.; Mira, J. (eds.): Brain Processes, Theories and Models. The MIT Press. pp. 24-36.
28. Newell, A. (1981): The Knowledge Level. AI Magazine, pp. 1-20.
29. Pylyshyn, Z. W. (1984): Computation and Cognition. MIT Press.
30. Rakic, P. (1975): Neuroscience Research Program Bulletin on L.C.N., 13, 3. Boston. pp. 299-314.
31. Rosenblueth, A., Wiener, N., and Bigelow, J. (1943): Behavior, Purpose and Teleology. Philosophy of Science 10.
32. Sánchez-Andrés, J.V.; Belmonte, C. (1995): La Metodología Experimental en Neurociencias: Representaciones Fragmentadas. En Barro, S. and Mira, J. (eds.): Computación Neuronal, Cap. III. Serv. de Pub. e Intercambio Científico. Univ. de Santiago de C. pp. 55-76.
33. Schwartz, J. (1987): Limits of IA. In: Shapiro, S.C. (ed.): Encyclopedia of Artificial Intelligence, Vol I. J. Wiley & Sons, N.Y. pp. 488-503.
34. Segev, I. (1992): Single Neurone Models: Oversimple, Complex and Reduced. Trends in Neurosciences, Vol. 15, No. 11, pp. 414-421.
35. Shepherd G.M.(ed). (1990): The Synaptic Organization of the Brain. Oxford Univ. Press.
36. Varela, F.J. (1979): Principles of Biological Autonomy. North-Holland. New York.
37. Wiener, N. (1947): Cybernetics. MIT Press and J. Wiley. New York.

(C) Consejo Superior de Investigaciones Científicas
Licencia Creative Commons 3.0 España (by-nc)
http://arbor.revistas.csic.es