Antonio Fernández-Caballero
Sergio Miguel Tomé (Eds.)

50 Años de la Inteligencia Artificial

XVI Escuela de Verano de Informática, EVI-2006
Albacete, España, 10-14 de Julio del 2006

Universidad de Castilla-La Mancha

Departamento de Sistemas Informáticos

Impreso en España. Printed in Spain.

ISBN 84-689-9328-X

Depósito Legal: AB-316-2006

Imprime: Gráficas Quintanilla. La Roda

Diseño de la cubierta: UGSC (Unidad de Gestión Sociocultural)

Presentación

El 31 de Agosto 1955, *J. McCarthy* (Dartmouth College, New Hampshire), *M.L. Minsky* (Harvard University), *N. Rochester* (I.B.M. Corporation) y C.E. Shannon (Bell Telephone Laboratories) lanzaron una propuesta para reunir en el verano de 1956 a un grupo de investigadores que quisieran trabajar sobre la conjetura de que cada aspecto del aprendizaje y cada característica de la inteligencia podían ser tan precisamente descritos que se podían crear máquinas que las simularan. El encuentro, celebrado en 1956 y ahora conocido como la conferencia de Dartmouth, se llevó a cabo con tal éxito que el evento acuñó el término *Inteligencia Artificial* y con él una nueva área científica de conocimiento. En el año 2006 se cumplen cincuenta años de la Conferencia de Dartmouth. Pero a pesar del tiempo transcurrido, el problema de encontrar las minuciosas descripciones de las características del cerebro y de la mente que fue mencionado en la propuesta de 1955 sigue tan vigente hoy, como ayer, a pesar del variado abanico de ciencias que lo abordan y estudian.

Albacete (España) ha sido en la semana del 10 al 14 de Julio la sede del evento internacional más importante en lengua castellana con el *Campus Multidisciplinar en Percepción e Inteligencia, CMPI-2006*. El Campus Multidisciplinar en Percepción e Inteligencia 2006 es un evento internacional en el que investigadores de diversas áreas relacionadas con la Percepción y la Inteligencia se encontrarán del 10 al 14 de Julio en el Campus Universitario de Albacete con el ánimo de recuperar el espíritu entusiasta de aquellos primeros días de la Inteligencia Artificial. En nuestra intención está el objetivo de crear un ambiente heterogéneo formado por especialistas de diversas áreas, cómo la Inteligencia Artificial, la Neurobiología, la Psicología, la Filosofía, la Lingüística, la Lógica, la Computación,, con el fin de intercambiar los conocimientos básicos de las diferentes áreas y de poner en contacto investigadores de los diferentes campos. El facilitar la creación de colaboraciones e investigaciones multidisciplinares es un objetivo prioritario de la propuesta.

La *Escuela de Verano en Percepción e Inteligencia: 50 Aniversario de la Inteligencia Artificial*, que ha dado lugar a esta publicación, se engloba como parte fundamental en el Campus Multidisciplinar sobre Percepción e Inteligencia. La XVI Escuela de Verano en Informática se ha ofertado en el seno de la XIX Edición de Cursos de Verano de la Universidad de Castilla-La Mancha. Pensada fundamentalmente para los alumnos de la Universidad de Castilla-La Mancha del Campus de Albacete, la Escuela de Verano sobre Percepción e Inteligencia ha cubierto aspectos de gran interés para las carreras de Informática, Medicina, Humanidades y Magisterio. Las clases magistrales de la Escuela de Verano han estado a cargo de importantes y reconocidos investigadores a nivel internacional. Todos ellos, así, y desde su propia experiencia, han proporcionado a los asistentes una visión muy clara del estado actual de las distintas ciencias que se ocupan de la Percepción y la Inteligencia.

Julio del 2006

Antonio Fernández-Caballero
Sergio Miguel Tomé
EVI-2006

Entidades Organizadoras

Universidad de Castilla-La Mancha
Departamento de Sistemas Informáticos, UCLM
Parque Científico y Tecnológico de Albacete
Excmo. Ayuntamiento de Albacete

Entidades Patrocinadoras

Ministerio de Educación y Ciencia
Junta de Comunidades de Castilla-La Mancha
(Consejería de Educación y Ciencia)
Caja Castilla-La Mancha
Excma. Diputación de Albacete

Índice general

La Inteligencia Artificial como Ciencia y como Ingeniería

José Mira Mira

Dpto. de Inteligencia Artificial, ETS Ing. Informática. UNED. Madrid. SPAIN
jmira@dia.uned.es

Abstract. En este trabajo, aprovechando la oportunidad del cincuenta aniversario de la Inteligencia Artificial (IA), reflexionamos sobre: (1) Las causas de la gran disparidad existente entre los objetivos iniciales de sintetizar inteligencia general en máquinas y los modestos resultados obtenidos tras medio siglo de trabajo y (2) Las distintas decisiones estratégicas que creemos que deben adoptarse para avanzar en el desarrollo de la IA como Ciencia Cognitiva y como Ingeniería del Conocimiento (IC).

1 Introducción

Actualmente se acepta, con alto grado de consenso, que el propósito general de la IA es desarrollar modelos conceptuales, procedimientos de reescritura formal de esos modelos, estrategias de programación y máquinas físicas para reproducir de la forma más eficiente y completa posible las tareas cognitivas y científico-técnicas más genuinas de los sistemas biológicos a los que hemos etiquetado de inteligentes [1]. También se acepta explícita o implícitamente que el avance de la IA está limitado por los avances en las técnicas de modelado, formalización y programación y por la evolución en los materiales y las arquitecturas de los computadores y los dispositivos electromecánicos ("robots") en los que se instala el cálculo. Esta definición conlleva la suposición subyacente de que efectivamente podemos sintetizar estas tareas cognitivas, de que podremos reducir el lenguaje natural a lenguajes formales, la semántica a la sintaxis, los símbolos neurofisiológicos a símbolos estáticos, el conocimiento a arquitecturas y, finalmente, el procesado biológico de la información a un cálculo.

La hipótesis fuerte de la IA fue que en un número corto de años (10, 20, 30, 40, ahora ya 50) iba a ser posible sintetizar los procesos cognitivos y conseguir "*inteligencia general en máquinas*" ("every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it" [2,3]. Es decir, que no existían impedimentos físicos, constitutivos ni formales para este objetivo y que sólo era cuestión de recursos. Cincuenta años después de la conferencia del Dartmouth College no todos los profesionales del campo estamos de acuerdo con esta afirmación, ni tampoco vemos que sea necesaria. La IA, como toda ciencia e ingeniería, debe de tener un carácter *instrumental* y sus objetivos

A. Fernández-Caballero & S. Miguel (Eds.): 50 Años de la Inteligencia Artificial, pp. 1-12, 2006.

no deben de ser otros que (1) ayudar a comprender los procesos neurofisiológicos, cognitivos y sociales y (2) prolongar los analizadores humanos y complementar sus deficiencias pero no necesariamente construir un humanoide no distinguible de los humanos mediante el test de Turing [4] o los experimentos conceptuales de Searle [5].

En nuestra opinión los objetivos iniciales de la IA fueron excesivos porque se ignoraron las enormes diferencias constitutivas entre el "conocer humano" y el conocimiento que los humanos hemos sido capaces de hacer residir en una máquina de cristal de silicio semiconductor gracias al desarrollo de arquitecturas, lenguajes y herramientas lógico-matemáticas que superponen organizaciones a la electrónica digital, de forma tal que a un observador humano le parece que la máquina "*es*" inteligente.

Curiosamente, este intento de añadir espectacularidad y nomenclatura cognitiva excesiva a nuestros programas y robots ha contribuido a oscurecer los sólidos resultados alcanzados por la computación, la robótica, la "visión" artificial, las técnicas de modelado conceptuad y formal, la estructuración del conocimiento de acuerdo con un plan estratégico para alcanzar una meta (diagnosticar, clasificar, predecir, etc.), las técnicas de representación formal (lógica, reglas, redes, marcos, objetos,...), los enlaces con biomateriales y las aplicaciones en dominios tan diversos y relevantes como la medicina, la meteorología, la economía, la industria, la educación o la WEB, por nombrar sólo los campos más conocidos.

El resto de este trabajo está estructurado de la siguiente forma. Primero distinguimos entre los objetivos de la IA como ciencia de los de la IA como Ingeniería del Conocimiento (IC) (sección 2). Después, en la sección tercera resumimos algunos puntos de la historia hasta llegar a la actitud sincrética actual en términos del uso combinado de distintos paradigmas básicos (simbólico, conexionista, situado, híbrido y bioinspirado). A continuación enumeraremos algunas de las posibles causas de la disparidad entre objetivos y resultados (sección 4) y proponemos un conjunto de sugerencias estratégicas y metodológicas que creemos que pueden ayudar a reducir esta disparidad y a reformular los objetivos de la IA y la IC de forma más modesta, clara, precisa e inequívoca que la usual (sección 5). En la última sección resumimos nuestras conclusiones.

2 La IA como Ciencia y como Ingeniería del Conocimiento

Aunque lo usual en IA e IC es mezclar conceptos cognitivos y del lenguaje natural (intención, propósito, ontología, semántica, emoción, memoria, aprendizaje,...) con otros computacionales (modelo, entidades y operadores lógico-matemáticos, autómatas, programas,...) suponiendo que estos conceptos tienen el mismo significado y las mismas funcionalidades en computación que en humanos, lo cierto es que no es así. Urge entonces distinguir entre IA como ciencia e IA como ingeniería (IC) para saber de qué hablamos cuando usamos estos conceptos (cuál es su referente) y evitar así equívocos. Esta mezcla de entidades cognitivas y abstractas y la asignación arbitraria de significados (el "no saber llevar bien la contabilidad", nos dice Maturana [6]) es

una de las causas fundamentales del problema de la disparidad entre objetivos y resultados. Prácticamente todo el lenguaje de la IA y la IC ha sido tomado de la biología en general y de la neurofisiología, la psicología cognitiva y la filosofía en particular. Recordemos de forma sucinta la distinción.

Entendida como ciencia, la fenomenología que aborda la IA engloba el conjunto de hechos asociados a la neurología y la cognición, desde los niveles subcelular y neuronal a los mecanismos y organizaciones de los que emergen las funciones globales de percepción, memoria, lenguaje, decisión, emoción y acción que han dado lugar a lo que llamamos *comportamiento inteligente en humanos*.

Lo que la IA busca, entendida como ciencia, es contribuir, junto con las otras ramas del conocimiento mencionadas previamente, a aproximar a la biología, la neurofisiología y la psicología al campo de las ciencias experimentales, tales como la física. Se busca entonces una *teoría* del ser y el pensar inteligente que, además, sea computable. Es decir, que sus modelos formales puedan ejecutarse en un sistema de cálculo y tener el mismo carácter predictivo que tienen, por ejemplo, las ecuaciones de Maxwell en el electromagnetismo. No es sorprendente entonces que con estos objetivos a largo plazo consideremos excesiva la conjetura fuerte de la IA.

Las ingenierías de la materia y la energía se basan en las sólidas teorías de la Física. Sin embargo la IC no puede basarse en una sólida *teoría del conocimiento* porque no disponemos de esa teoría. Esta es otra de las razones de la disparidad ostensible entre objetivos y resultados de la IA. Por consiguiente, parece razonable dejar el tiempo necesario para que la parte teórica de la IA contribuya a obtener una teoría computable del conocer humano y, mientras tanto, redefinir los objetivos de la IC de forma más modesta, teniendo en cuenta el carácter limitado, incompleto y poco preciso del conocimiento del que disponemos sobre dos tipos de tareas: (1) Tareas básicas e inespecíficas usuales en humanos, independientemente de su actividad profesional, tales como ver, oír, interpretar el medio, planificar, aprender, controlar las acciones encaminadas a moverse y manipular un medio, etc. y (2) Tareas científico-técnicas en dominios estrechos (diagnosticar en medicina, configurar y diseñar sistemas, etc...). Es crucial aceptar inicialmente las limitaciones en el alcance, las funcionalidades y la autonomía de estos sistemas de IA asociadas al desconocimiento de la lógica de la cognición y a las diferencias constitutivas entre el cuerpo biológico y el robot, entre el lenguaje formal y el lenguaje natural, entre la sintaxis y la semántica.

En la mayoría de los desarrollos de la IC se procede de acuerdo con los siguientes pasos: (1) Se parte de una descripción en lenguaje natural de las interacciones de un humano con el entorno en el que se desarrolla la tarea que queremos sintetizar (figura 1). Es decir, del método usado por el experto humano para “resolver” esa tarea. (2) Después se modela esta descripción usando diferentes meta-modelos a los que llamamos paradigmas (simbólico o representacional, conexionista, situado o híbrido). Cada una de estas formas de modelado conceptual es esencialmente un procedimiento de descomposición de la tarea en subtareas hasta llegar al nivel de inferencias primitivas que son aquellas componentes del razonamiento que ya no necesitan una descomposición posterior (“seleccionar”, “comparar”,…) porque ya se pueden implementar usando sólo conocimiento del dominio. (3) El tercer paso es la reescritura

formal de las inferencias y los "roles" estáticos y dinámicos de acuerdo con el paradigma elegido que, a su vez, es consecuencia del balance entre los datos y el conocimiento disponible y del tipo de datos (etiquetados o no etiquetados) y de conocimiento (asociado a los sensores y efectores del mecanismo en el que reside o para ser usado por un operador humano). (4) Finalmente, se programan los operadores. Para aquellas tareas en las que el interfaz es físico y no humano (sensores y efectores de un robot concreto, por ejemplo), es imprescindible implementar también el conocimiento asociado al "cuerpo" soporte del cálculo, marcando así un límite claro a las funcionalidades del sistema.

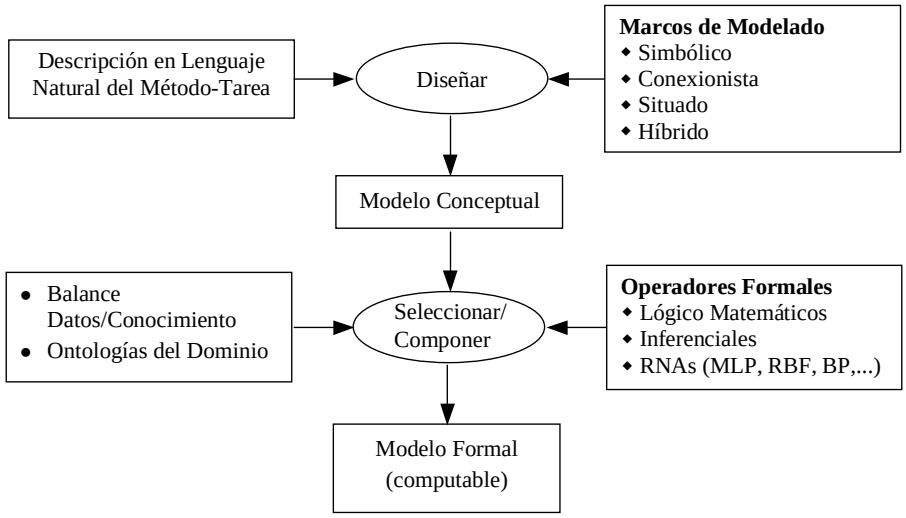


Fig. 1. Esquema cualitativo de la IC Se modela a partir de descripciones en lenguaje natural del procedimiento usado por un humano para resolver una tarea.

3 Perspectiva Histórica

El sueño de mecanizar los procesos del pensamiento (lo que ahora llamamos "hacer computacional el conocimiento humano" o sintetizar sus procesos cognoscitivos en sistemas de IA) es muy antiguo y procede –como casi todo– de los griegos. Aunque se suele reconocer que el nacimiento de la IA debe asociarse a la conferencia que se organizó en 1956 en el Dartmouth College, su nacimiento real pasa por Platón, Descartes, Boole, Leibnitz y Hobbes, encuentra una etapa neurocibernética a partir de 1943 con los trabajos pioneros de W.S. McCulloch, W. Pitts [7], K. Craik [8], J. von Neumann, N. Wiener [9,10], J. McCarthy, A. Newell, M.L. Minsky, S.C. Kleene, M. D. Davis, A.M. Uttley y C. Shannon [11] y entra después en una etapa lógica [12] de

dominios formales (micromundos) y de búsqueda heurística en un espacio de estados [13].

Hacia 1975 se da un cambio brusco en la orientación de los trabajos en IA haciendo énfasis en la importancia del conocimiento frente a los mecanismos generales de inferencia. Así, se abandona la búsqueda de "solucionadores generales de problemas" y se focaliza el trabajo en el modelado del conocimiento asociado a los métodos usados por los expertos humanos para resolver problemas en dominios estrechos, dando lugar al campo de los *sistemas basados en conocimiento* (los SBC) [14].

En torno a 1986 la IA mira de nuevo a la neurocibernética y se produce un fuerte renacimiento del conexionismo con el desarrollo de redes de "neuronas" artificiales y algoritmos de aprendizaje supervisado y no supervisado que sustituyen parcialmente a la programación en aquellas aplicaciones en las que se dispone de más datos que conocimiento [15]. También crecen en esta época las aproximaciones probabilísticas (bayesianas) y posibilísticas (borrosas) al tratamiento de la incertidumbre.

En la misma época y con aportaciones procedentes especialmente del campo de la robótica, hay una nueva aproximación a la IA llamada "situada" o reactiva, basada en la observación etológica de la conducta a imitar y liderada por investigadores seguidores de Brooks [16], Arkin [17] y Clancey [18], entre otros. La clave de esta aproximación es su esfuerzo por no separar la inteligencia del cuerpo que la soporta. W.S. McCulloch llamaba a este problema los "*Embodiments of Mind*" [19].

Los números 1 y 2 de 1993 del volumen 59 de la revista *Artificial Intelligence*, reeditado por Bobrow en 1994 [20] son representativos del estado de conocimiento en esa época. Incluyen trabajos de McCarthy, Moore, Newell, Nilsson, Gordon y Shortlife, Pearl, Simon, Kuipers, Aikins, Arbib y Minsky entre otros autores relevantes del campo.

La situación actual se caracteriza por (1) Un fuerte movimiento en torno a la reusabilidad de componentes en el modelado del conocimiento con el desarrollo de bibliotecas de tareas, métodos, inferencias y roles y con el interés por las ontologías y los servidores de terminología unificada [21], (2) La aceptación de la naturaleza híbrida de la mayoría de los problemas y la consiguiente necesidad de usar arquitecturas híbridas [22] en su solución (con elementos simbólicos, conexionistas, situados y de tratamiento de la incertidumbre –bayesianos o borrosos–) y (3) un fuerte movimiento interdisciplinario en el que de nuevo se mira a la biología con un doble propósito (conseguir inspiración para nuevos modelos de cálculo y usar el material biológico como alternativa al cristal de silicio semiconductor para atravesar la frontera de la nanotecnología) [23,24]. En este movimiento se enmarcan aproximaciones tales como los algoritmos genéticos y la programación evolutiva, la computación celular y molecular (DNA) y la computación cuántica [25], siguiendo las propuestas de Holland, Feyman, Adleman, Bennett y muchos otros autores que se agrupan bajo el paraguas de la llamada computación natural.

Finalmente, para completar la visión histórica de estos cincuenta años de la IA e IC es necesario recordar dos aspectos metodológicos que considero de especial interés. El primero es la introducción del *nivel de conocimiento* por Newell [26] y Marr [27]. Se reconoce así que no basta con conocer el nivel físico (la máquina) y el nivel de los símbolos (el programa) para saber realmente de lo que trata un cálculo. Es

necesario conocer también la teoría de ese cálculo, su modelo conceptual, las tablas de correspondencias con el modelo formal y la semántica externa, ya que al computador sólo pasa el *modelo formal subyacente* a todas estas descripciones [22,28], El segundo avance metodológico es la introducción de la figura del *observador externo* en la descripción de los programas de IA y de los procesos neurofisiológicos. Esto nos permite distinguir en cada uno de los tres niveles de descripción (máquinas, símbolos y conocimiento) dos *dominios de causalidad*, el propio de cada nivel (causalidad interna) y el del observador (causalidad externa). Esta figura del observador procede de la física, fue introducida en biología por Maturana [6,29] y Varela [30] y en computación por Mira y Delgado [31,32,28,22]. Si a la hora de valorar las potenciales funcionalidades se tuviera clara la distinción entre lo que pertenece al dominio propio (es decir lo que reside en la máquina) y lo que pertenece al dominio del observador (es decir las etiquetas lingüísticas y la semántica) sería evidente también la distinción entre el nivel de inteligencia que realmente hemos sido capaces de computar y aquellos otros componentes de la inteligencia humana que sólo existen en la mente del intérprete.

4 Fronteras entre el conocer humano y el conocer de las máquinas

Al reflexionar sobre los logros de la IA y la IC en estos cincuenta años nos parece evidente que se ha avanzado razonablemente en su perspectiva aplicada, Sin embargo el objetivo fundamental de 1956, "*sintetizar inteligencia general en máquinas*", está todavía muy lejos [33]. Entre las posibles razones de esta distancia nos parece que se encuentran las siguientes [34]:

4.1 Las grandes diferencias constitutivas entre los sistemas biológicos y las máquinas

Si todo conocer depende de los materiales y los mecanismos constituyentes del sistema que conoce, es claro que el conocer humano es distinto del conocer propio de las máquinas. Ambos dependen de la fenomenología que sus elementos constituyentes *generan al operar*. Resumimos aquí estas diferencias en los tres niveles de descripción de un cálculo. En el nivel físico de las máquinas las entidades constituyentes son circuitos lógicos y retardos que sólo permiten establecer distinciones binarias (0,1) sobre expresiones lógicas y transiciones de estado en autómatas finitos. Todo el resto del conocimiento que puede acomodar este nivel está asociado a la arquitectura (al lenguaje máquina). El cuerpo del computador es de cristal semiconductor, con arquitectura fija, estática y con semántica impuesta. Por el contrario, las entidades constituyentes del tejido nervioso (proteínas, canales iónicos, neuronas) permiten acomodar todo el conocimiento que la genética, la evolución, la historia y la cultura han aportado al soporte neurofisiológico de la inteligencia. El soporte neurofisiológico del "cálculo inteligente" es autónomo, autopoyético, dinámico, adaptivo, con

semántica emergente y con una arquitectura siempre inacabada y, por consiguiente, finalmente única e irrepetible.

También en el nivel de los símbolos hay diferencias esenciales. En los computadores programables convencionales usamos símbolos "fríos", descriptivos, estáticos, de semántica arbitraria de acuerdo con la propuesta de Newell y Simon en su "Physical Symbol System Hypothesis" [35]. Por el contrario, el símbolo neurofisiológico es dinámico y está asociado a los mecanismos que lo generan e interpretan [36]. Sin el conocimiento de la historia evolutiva de un animal concreto (la rana) y de su medio (la charca) será difícil entender la representación simbólica que construyen sus células ganglionares (las "*bug detectors*"), por ejemplo.

Finalmente, las diferencias en el nivel de conocimiento son las de observación más directa, tanto por introspección como por consideración del cuerpo de conocimientos acumulados por la fisiología, la psicología, la lingüística, la sociología y la filosofía. La inteligencia humana está asociada al lenguaje natural, a los propósitos, intenciones, motivaciones y emociones. Si tiene sentido hablar aquí de cálculo este cálculo es *semántico* e *intencional*. Por el contrario ya hemos recordado que al computador sólo pasa el "modelo formal subyacente" asociado a un lenguaje formal. Es decir todo el cálculo actual es *sintáctico* y en *extenso*. Hoy por hoy no tenemos una idea razonablemente clara, consensuada y físicamente realizable de cómo podríamos construir un cálculo intencional.

4.2 Ignorancia sobre la Lógica de los procesos cognitivos y falta de teoría

La segunda causa de la disparidad es el desconocimiento que tenemos de la fisiología y la lógica de los procesos cognitivos y la consiguiente dificultad de reproducir algo que no se conoce. La inteligencia natural es un concepto muy amplio que engloba un gran número de habilidades, más allá de la mera solución de problemas científico-técnicos en dominios estrechos.

Asociada a esta ignorancia ha estado la prisa excesiva en hacer ingeniería (desarrollar aplicaciones) sin disponer del soporte científico previo y necesario: una teoría bien fundada experimentalmente, invariante y con capacidad predictiva sobre el conocer humano. Esto ha llevado a una cierta superficialidad en las propuestas de la IC que han enmascarado la falta de teoría con un uso inadecuado del lenguaje. Muchas de las propuestas de la IA y la IC sólo existen en el lenguaje del observador. El uso no justificado de términos cognitivos, dando por supuesto que tienen el mismo significado en humanos que en computación ha contribuido enormemente a crear el espejismo de que el nombre de las etiquetas que usa el programador se corresponde exactamente con lo que el computador hace con la entidad abstracta subyacente a esa etiqueta. De hecho es al revés, basta con intentar hacer ingeniería inversa de un programa concreto para comprobar que hay muchas descripciones compatibles con el mismo. También hay muchos significados compatibles con los símbolos usados en la formulación de una ecuación diferencial (campo eléctrico, magnético, gravitatorio, número de elementos de una población, etc.) y no por eso adscribimos las semánticas propias del electromagnetismo, la gravitación, la mecánica cuántica o la sociología

poblacional a la semántica propia de las entidades abstractas de un texto sobre ecuaciones diferenciales (variables, conjuntos, sumas, restas, derivadas, integrales,…), tal como se estudia en una Facultad de Matemáticas. La Facultad de Física es "otro edificio" que está al lado. Las Facultades de Sociología y Psicología también están en otros edificios próximos, pero no son el mismo que el de la Lógica y las Matemáticas.

4.3 Falta de herramientas formales y de nuevos modelos de computación

En todos los desarrollos de IA e IC, al intentar reescribir formalmente los modelos conceptuales, sólo disponemos de las matemáticas heredadas de la física (el álgebra y el cálculo), la lógica y la teoría de autómatas, junto con algunos elementos de cálculo de probabilidades y estadística. La duda es si estas herramientas formales son o no suficientes para describir los procesos cognitivos. En nuestra opinión, hacen falta nuevas herramientas formales para captar el carácter dinámico, adaptivo, impreciso, contradictorio, intencional y semántico del conocimiento humano. Algo análogo creemos que pasa también con los fundamentos de la computación (máquinas de Turing, lenguajes de Chomsky o redes de neuronas formales en el sentido de McCulloch-Pitts) y la arquitectura von Neuman, que es posible que no sean suficientes (quizás ni siquiera las más adecuadas) para modelar los procesos mentales. Quizás sea posible explicar el pensamiento sin computación. Interpretando a Bronowski, si queremos modelar los procesos soporte de la inteligencia en términos computables, tendremos que reducirlos a un lenguaje formal y no está claro que la naturaleza de lo vivo permita esa reducción. Los nuevos desarrollos interdisciplinarios en la frontera entre biología y computación tienen a su Gödel en Bronowski.

5 Algunas Sugerencias Estratégicas y Metodológicas

Para contribuir a disminuir la disparidad entre objetivos y resultados en IA e IC y hacerlas más robustas, presentamos de forma resumida algunas sugerencias. La primera sugerencia es aumentar el esfuerzo dedicado a los fundamentos de la IA, a la construcción de una teoría computable del conocimiento humano y al desarrollo de herramientas conceptuales y formales más adecuadas para describir los procesos mentales, específicos de los seres vivos.

La segunda sugerencia estratégica tiene que ver con la *distinción* clara entre el concepto de inteligencia general en humanos y los objetivos realizables de la IC. Aquí, en el contexto de la IC, "inteligencia Artificial" se entiende como el conjunto de tareas, métodos, herramientas formales, estrategias de programación y mecanismos físicos usados para la automatización del proceso de "solución de un conjunto de problemas científico-técnicos", basadas en el uso intensivo de conocimiento humano, en general no analítico. Aún con esta limitación, el objetivo a medio plazo de la IC es suficientemente ambicioso.

Una componente importante de esta segunda sugerencia tiene que ver con la conveniencia de usar un lenguaje más modesto que el cognitivo y más próximo a la semántica de las entidades que constituyen un cálculo. Es decir, más próximo a los conceptos formales, a la teoría de autómatas y a los lenguajes de programación que a los conceptos del lenguaje natural, tal como lo usan los humanos para describir los procesos cognitivos, para comunicarse entre sí e, incluso, para modelar y describir un cálculo en el dominio del observador externo, sin exigencia de contrapartida causal en la máquina física, como yo estoy haciendo aquí ahora. Hay que "rebajar" la nomenclatura procurando no sobre-nombrar las cosas. Así, proponemos *reformular* los objetivos de la IC usando sólo términos que tengan un significado computacional claro e inequívoco (tareas, métodos, inferencias, roles, entidades y relaciones del dominio, operadores, autómatas, lógica, etc...), sin la intención, o suposición implícita, de que estas componentes de modelado y formalización (entidades abstractas) tengan que ser la contrapartida computable de las componentes del lenguaje natural (nombres-conceptos, verbos-inferencias, condicionales,...) que pretenden modelar. El propósito de la IC es el modelado conceptual y la reescritura formal del procedimiento de solución de un problema de forma tal que sea computable con las restricciones inexcusables de las herramientas formales y de las máquinas físicas de las que disponemos en la actualidad. El trabajo real está ahora en encontrar procedimientos de modelar conocimiento y de reescribir formalmente esos modelos que hagan posible (con la mínima pérdida de semántica) el enlace con las organizaciones superpuestas a la arquitectura de la máquina para las que ya existen traductores.

Para llevar a cabo esta reformulación de los objetivos y temas abordados por la IC proponemos la tarea metodológica de "*desmontar*" los diferentes ítems que aparecen actualmente bajo el paraguas de IA o IC (computational intelligence, soft-computing, data mining, neural networks, computing with words, genetic algorithms, case-based reasoning, models of users and students, behaviors, perceptions, purposes, artificial life, knowledge discovery, intelligent agents,...), en términos de sus componentes de modelado y formalización constituyentes (reglas, clasificaciones numéricas, planificación, búsqueda, ajuste de parámetros, ...). Posteriormente, será posible *organizar* estas entidades en bibliotecas de "*componentes reutilizables*" de modelado y formalización y *proponer* procedimientos de *selección* y *ensamblado* de esas componentes de acuerdo con el balance entre datos y conocimiento disponibles en cada aplicación concreta.

La tercera decisión estratégica aconsejada es promover la *interacción* entre las ciencias de lo vivo y la computación, en el clásico estilo de la época fundacional de la cibernética (1943-1950), previa a la visión simbólica de la IA (1956). Hemos empezado diciendo que una de las causas de la disparidad entre objetivos y los resultados de la IC era la mezcla y confusión entre conceptos biológicos, psicológicos y fisiológicos por una parte y conceptos puramente computacionales o formales por otra, junto con la falta de reconocimiento de las enormes diferencias constitutivas entre los sistemas biológicos y las máquinas. Sin embargo, una vez que queda clara la distinción, creemos que es útil mirar a la naturaleza y a las ciencias de lo vivo con un doble objetivo. Por una parte como fuente de inspiración para encontrar nuevos ma-

teriales como potencial soporte del cálculo (DNA, membranas, moléculas) y para formular nuevos mecanismos, nuevas estrategias de programación y nuevos modelos de utilidad en IC. Esta orientación se conoce con el nombre de *computación e ingeniería bioinspirada*. Por otro lado, es cada vez más necesario el trabajo interdisciplinario en el que la física, las ingenierías, la lógica, las matemáticas y, resumiéndolas a todas, la *computación*, ayuden a las ciencias de lo vivo a experimentar y formular teorías explicativas que cierren el lazo de realimentación proporcionando a la IA y la IC el fundamento científico que tanto necesitan. Esta orientación complementaria se conoce con el nombre de *Neurociencia Computacional* y, en un contexto más amplio, como *visión computacional* de las *ciencias de lo vivo*. Para los lectores de cierta edad estamos hablando de *Biocibernética* y *Biónica*.

6 Conclusiones

Hemos reflexionado sobre la disparidad entre los objetivos y los resultados de la IA y la IC y sobre las posibles causas de esa disparidad. Finalmente hemos propuesto algunas sugerencias para contribuir al desarrollo de una IA tan robusta como la física y una IC tan robusta como las actuales ingenierías de la materia y la energía. Queremos creer que con este análisis hemos contribuido al menos a eliminar el error más frecuente en estos 50 años de desarrollo de la IA consistente en considerar que las palabras que usamos en la construcción de nuestros modelos conceptuales son directamente computables, que la descripción de un comportamiento coincide con el mecanismo del que emerge ese comportamiento que el "mapa es lo mismo que el territorio" (Korzybski). Este error histórico ha dado pie a suponer que nuestros programas de IA acomodan mucho más conocimiento del que realmente acomodan. Lo cierto es que gran parte del conocimiento supuestamente computable queda fuera del computador, y aquí está el trabajo real, en conseguir que nuevas capas organizativas intermedias (nuevos modelos y mecanismos de formalización) permitan que cada vez sea mayor la cantidad de conocimiento residente en el sistema físico, lógico y conceptual al que en el futuro llamemos "*nuevo computador bioinspirado*", o si se quiere, nuevo sistema de IA.

Actualmente, la cuestión básica en la perspectiva aplicada de la IA no es lo que "pueden o no pueden hacer los computadores" [33], sino la cantidad y tipo de conocimientos que los humanos seremos capaces de modelar, formalizar y programar de forma tal que finalmente sea computable en una máquina. Esta es una cuestión abierta cuya respuesta depende de la evolución de los materiales y arquitecturas y de los avances en las técnicas de modelado, formalización y programación. En cualquier caso, al plantear de esta forma la cuestión, la situamos en el contexto serio de la ciencia y las ingenierías convencionales, sin excesos de nomenclatura ni significados no justificables y, por consiguiente, contribuimos a acotar el espacio de búsqueda de soluciones.

Finalmente, hay todavía muchas "cuestiones básicas" abiertas en la perspectiva científica de la IA. Un primer paso en su solución es reconocer con modestia la enorme dificultad asociada a intentar convertir a las "ciencias" de lo humano en una ciencia equiparable a la Física. El futuro es prometedor y apasionante. La doble aventura pluridisciplinar de "*conocer el conocer*" e "*ingenierizar lo conocido*" es un desafío que invita a los pioneros humanistas, científicos e ingenieros de mente amplia e inquieta. De momento celebremos los 50 primeros años de esta doble aventura.

Referencias

1. [Mira, Delgado, 95] Mira, J., Delgado, A.E.: Perspectiva Histórica Conceptual. En Aspectos Básicos de la Inteligencia Artificial. Sanz y Torres, Madrid (1995) 1-51
2. [McCarthy, Minsky, Richester, Shannon, 55] McCarthy, J., Minsky, M.L., Richester, N., Shannon, C.E.: A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Int. report. Hannover, New Hampshire (1955)
3. [Rich, 90] Rich, E.: Artificial Intelligence. In Shapiro, S.C. (Ed.): Encyclopaedia of Artificial Intelligence Vol. I. Wiley& Sons, N. York (1990) 9-16
4. [Moor, 87] Moor, J.H.: Turing Test. In: Shapiro, S.C. (Ed.): Encyclopaedia of Artificial Intelligence Vol II. J. Wiley & Sons, New York (1987) 1126-1130
5. [Searle, 85] Searle, J.: Minds, Brains, and Programs. In Haugeland, J. (Ed.): Mind Design. Philosophy, Psychology, Artificial Intelligence. The MIT Press (1985) 282-306
6. [Maturana, 75] Maturana, H.R.: The Organization of the Living: A theory of the Living Organization. Int. J. Man-Machine Studies **7** (1975) 313-332.
7. [McCulloch & Pitts, 43] McCulloch, W.S., Pitts, W.: A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics Vol. 5 Chicago Univ. Press (1943) 115-133.
8. [Craik, 43] Craik, K.: The Nature of Explanation. Cambridge University Press, Cambridge (1943)
9. [Rosenblueth et al, 43] Rosenblueth, A., Wiener, N., Bigelow, J.: Behavior, Purpose and Teleology. Philosophy of Science 10 (1943)
10. [Wiener, 47] Wiener, N.: Cybernetics. The Technology Press. J. Wiley & Sons, New York (1947)
11. [Shannon & McCarthy eds., 56] Shannon, C.E., McCarthy, J. (Eds.): Automata Studies. Princeton University Press, N. Jersey (1956)
12. [Newell & Simon, 56] Newell, A., Simon, H.A.: The Logic Theory Machine. IRE Trans. Information Theory **2** (1956) 61-79.
13. [Barr & Feigenbaum, 81] Barr, A., Feigenbaum, E.A.: The Handbook of Artificial Intelligence, Vol. 1 and II. William Kaufmann (1981)
14. [Buchanan, Feigenbaum, 78] Buchanan, B.G., Feigenbaum, E.A.: DENDRAL and Meta-DENDRAL. Artificial Intelligence 11 (1978) 5-24
15. [Haykin, 99] Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice-Hall (1999)
16. [Brooks, 91] Brooks, R.A.: Intelligence without Reason. MIT A.I. Memo Nº. 1293 (1991)
17. [Arkin, 98] Arkin, R.C.: Behavior-based Robotics. The MIT Press (1998)
18. [Clancey, 97] Clancey, W.J.: Situated Cognition. On human knowledge and computer representations. Univ. Press, Cambridge (1997)

19. [McCulloch, 65] McCulloch, W.S.: Embodiments of Mind. The MIT Press. Cambridge, Mass (1965)
20. [Bobrow, 94] Bobrow, D.G. (Ed.): Artificial Intelligence in Perspective. The MIT Press (1994)
21. [Schreiber, 99] Schreiber, et al.: Engineering and Managing Knowledge: The Common-KADS methodology. The MIT Press, Mass(1999)
22. [Mira, Delgado, 03] Mira, J., Delgado, A.E.: Where is knowledge in robotics? Some methodological Issues on symbolic and connectionist perspectives of AI. In Zhou Ch, Maravall D, Da Rua (Eds.): Autonomous Robotic Systems. Physical-Verlag. Springer-Verlag, Berlin (2003) 3–34
23. [Mira, 05] Mira, J.: On the Use of the Computational Paradigm in Neurophysiology and Cognitive Science. In José Mira and José R. Álvarez (Eds.):. Mechanisms, Symbols, and Models Underlying Cognition. IWINAC 2005, LNCS 3561, pp. 1-15. Sprnger, 2005.
24. [Barro, Bugarín (eds.) 02] Barro, S., Bugarín, A.J. (eds.): Fronteras de la Computación. Dintel y Díaz de Santos, Santiago de Compostela (2002)
25. [Nielsen, Chuang, 01] Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press (2001)
26. [Newell, 81] Newell, A.: The knowledge level. AI Magazine Vol. 120 (1981)
27. [Marr, 82] Marr, D.: Visión. Freeman, New York (1982)
28. [Mira, Delgado, 95] Mira, J., Delgado, A.E.: Aspectos Metodológicos en IA. En Aspectos Básicos de la Inteligencia Artificial, Cap. II. Sanz y Torres. Madrid (1995) 53-87
29. [Maturana, 02] Maturana, H.R.: Ontology of Observing. The Biological Foundations of Self Consciousness and the Pphysical Domain Existence. http://www.inteco.cl/biology/ontology/ (2002)
30. [Varela, 79] Varela, F.J.: Principles of Biological Autonomy. The North Holland Series in General Systems Research. North-Holland, New York (1979)
31. [Mira, Delgado, 87] Mira, J., Delgado, A.E.: Some Comments on the Antropocentric Viewpoint in the Neurocybernetic Methodology. Proc of the Seventh International Congress of Cybernetics and Systems 2 (1987) 891-95.
32. [Mira, Delgado, 03] Mira, J., Delgado, A.E.: Neural Modeling in Cerebral Dynamics. BioSystems 71, (2003) 133-144
33. [Dreyfus, 94] Dreyfus, H.L.: What Computers Still Can't do. The MIT Press, Camb. Mass (1994)
34. [Mira, 05] Mira, J.: On the Physical Formal and Semantic Frontiers Between Human Knowing and Machine Knowing. In R. MorenoDíaz, F.Pichler, A. Quesada Arencibia (Eds.): Computer Aided Systems Theory. LNCS Vol. 3643. Springer-Verlag (2005) 1-8
35. [Newell, Simon, 76] Newell, A., Simon, H.A.: Computer Science as Empirical Inquiry: Symbols and Search. Communications of ACM **19** (1976) 113-126.
36. [Mira, 05] Mira, J.: On the Use of the Computational Paradigm in Neurophysiology and Cognitive Science". In J. Mira and J.R. Álvarez (Eds.): Mechanisms, Symbols, and Models Underlying Cognition. LNCS 3561., Springer-Verlag (2005) 1-15

Lógica natural e Inteligencia artificial

Gladys Palau

Universidad de Buenos Aires- Universidad Nacional
de La Plata, Argentina

I

Comenzaré esta disertación tratando de delimitar lo que habré de entender por "lógica natural", o si se prefiere "lógica del sentido común" o "*folk* a los efectos de diferenciarla de la lógica en tanto ciencia y de esclarecer sus relaciones con su nueva pariente, la Inteligencia Artificial (I.A.).

Hoy en día hay acuerdo de entender por lógica clásica (LC) a la disciplina que, originada en Aristóteles, se plasmó en tanto conjunto de teoremas o verdades lógicas, en las presentaciones axiomáticas de G. Frege y B. Russell y, en tanto conjunto de reglas de inferencia, en los cálculos de deducción natural y de secuentes de G. Gentzen. También se acuerda en considerar lógica clásica a todas las extensiones conservativas de LC, tales como la familia de las lógicas modales construida a partir de los sistemas de I. C. Lewis. Por otra parte, cabe señalar que los sistemas de lógica intuicionista, de lógica de la relevancia, los multivalentes, y en general, los sistemas lógicos divergentes, constituyen lógicas subclásicas, ya que rechazan algunos principios de la lógica clásica, sin por ello abandonar las propiedades esenciales de la noción de consecuencia lógica clásica, a saber, Reflexividad, Monotonía y Corte. Tanto la lógica clásica como las lógicas subclásicas tienen las siguientes dos propiedades que interesa destacar a nuestros fines: 1) son lógicas deductivas o preservativas de la verdad, es decir, su noción de consecuencia lógica satisface Monotonía y 2) son formales ya que en todas ellas vale la regla de Sustitución Uniforme.

Previamente a la caracterización que nos hemos propuesto, creemos ilustrativo e interesante decir algunas palabras sobre la lógica aristotélica y la lógica del discurso de sentido común. Es sabido que la obra de Aristóteles es presentada siempre como iniciadora "*ex nihilo*" de la lógica formal. En cierto sentido esta afirmación es correcta, ya que su teoría sobre el silogismo y su caracterización de ciencia demostrativa constituyeron las primeras reflexiones teórico-sistemáticas sobre el objeto de la lógica y el concepto de demostración. Sin embargo, en su obra *The Development of Logic* (1962), William y Martha Kneale sostienen que para poder haber reflexionado sobre los principios de la validez, Aristóteles debió contar con una suma considerable de argumentaciones y material inferencial sobre el cual realizó

A. Fernández-Caballero & S. Miguel (Eds.): 50 Años de la Inteligencia Artificial, pp 13- 24, 2006.

tales reflexiones, y que en nuestra terminología bien podríamos calificar como lógica del sentido común o lógica natural. Estos serían: en primer lugar, los razonamientos empleados en las demostraciones matemáticas; en segundo lugar, la argumentación dialéctica con fines metafísicos[1] y en tercer lugar, las argumentaciones jurídicas y los discursos de índole práctica, de los que parece extrajo los *sofismas*.

En sintonía más bien con esta idea, en un artículo de 1982, el lógico sudamericano Francisco Miró Quesada[2], afirma:

(...) *hay una lógica que es la que usamos cuando pensamos racionalmente de manera espontánea y ella es el fundamento en torno del cual se constituyen todas las lógicas, es la base sin la que ninguna lógica formal, por más aberrante que sea, tiene sentido.*

Siguiendo una denominación de R. Meyer en su artículo "Why I am not a relevantist", Miró Quesada bautiza a esta lógica con el nombre *Nuestra Lógica*. Aunque sus límites sean imprecisos y su definición no sea rigurosa, la proponemos como equivalente a lo que comúnmente hoy en día se entiende como *lógica natural* o *lógica del sentido común.*

Ejemplos que evidencien la existencia de esta lógica no faltan. Pensemos simplemente en la práctica muy común de los lógicos cuando discuten acerca de la legitimación o rechazo de algún principio lógico o de alguna regla lógica. Pensemos, en los argumentos que exponían Filón y Diodoro cuando discutían sobre las condiciones de verdad del hoy llamado condicional material; o cuando el mismo Frege, analizando la misma constante lógica, se preguntaba por el único caso en el que era justo refutar un enunciado condicional matemático; o cuando C.I. Lewis, frente a las "perplejidades" que generaba al sentido común el condicional material, tales como que una proposición verdadera es implicada por cualquier proposición o que una proposición falsa implica cualquier proposición, propuso introducir el condicional estricto; o, finalmente, cuando se constató que el condicional material convertía en enunciados verdaderos a todos los enunciados contrafácticos, muchos de los cuales eran naturalmente falsos, tal como "Si esta figura geométrica no hubiera tenido tres lados, sería un cuadrado". Si se analizan estos ejemplos a la luz de las argumentaciones que los soportan, se constata que todos ellos acuden a una suerte de "intuiciones" lógicas que no son otra cosa que cierto tipo de normatividad intrínseca a la lógica natural. De hecho, estas "intuiciones" juegan un doble papel: por un lado son la base que origina o motiva cualquier sistema lógico, y por el otro, constituyen la base a la que se recurre para verificar la adecuación del sistema lógico construido. En el primer caso, los sistemas lógicos consisten en reconstrucciones formales de

[1]Expuesta principalmente en los diálogos de Platón, de donde Aristóteles parece haber extraído el llamado *argumento refutatorio* (Si P entonces Q, pero no-Q, luego no-P)y la *reductio ad impossibile* empleada por Zenón;

[2] "Nuestra lógica", Revista Sudamericana de filosofía, III,1,pp. 3-13

determinadas intuiciones y en el segundo, ellas se convierten en una especie de "prueba empírica" a fin de mostrar la adecuación o inadecuación del sistema lógico en cuestión. En síntesis, este doble juego tan común hoy en día en la práctica lógica, implícitamente presupone la existencia de una lógica natural, previa a cualquier lógica formal y que se la supone dotada de una normatividad que, como mostraremos más adelante, oficia como control para admitir, rechazar o adecuar (o "ajustar") una determinada inferencia en un sistema formal.

Desde el campo de las ciencias cognitivas, diversas investigaciones también avalan la existencia de una lógica natural. En efecto, aunque existen enfoques constructivistas diferentes, todos ellos coinciden en reconocer que cualquier ser humano está habilitado para razonar. Es decir, que el ser humano naturalmente posee una *competencia lógica* y explican además las capacidades y habilidades lógicas de toda persona en términos de *competencia* y *ejecución* (*perfomance*). Sin embargo, el problema que más profundamente divide a los investigadores en esta área concierne al papel que se atribuye a los sistemas lógicos en la descripción de tal competencia lógica.

Por ejemplo, M. Braine sostiene que existe una lógica natural compuesta por un conjunto de 14 inferencias lógicas naturales básicas, entre las que se encuentran. *Conjunción, Simplificación, Doble Negación, Silogismo Disyuntivo, Dilemas Constructivos Simple y Compuesto* y formas especiales del *Teorema de la Deducción* y del *Reductio ad Absurdum.* Estas se constituyen en la práctica del razonamiento diario, y a partir ellas se construyen otras cadenas cortas de inferencias, restringidas por el contexto pragmático en el que se ejecutan.

Otros, como E. K. Scholnik, sostienen que la suposición de reglas básicas, tales como los esquemas inferenciales de Braine es demasiado fuerte, pero que, sin embargo, hay procesos lógicos no explicitados, tales como la inferencia por el absurdo usado en la búsqueda de los contraejemplos. También cabe citar la propuesta de J. Piaget y sus seguidores tales como Grize, Matalon, Le Bonniec, Overton y R. García, según la cual hay una lógica natural que se constituye a partir de las acciones de las acciones del sujeto con el mundo desde su niñez, cuya estructura es caracterizada por medios algebraicos.[3]

Contrariamente a este enfoque, Johnson-Laird en diversos trabajos ha afirmado categóricamente que las inferencias de la lógica natural no son deductivas y por ello no pueden ser expresadas por reglas de inferencia formales. Más aún, si hubiera reglas, ellas no serían las reglas de inferencia de la lógica formal clásica ya que, en la vida real, los hombres a menudo carecen de la información completa como para realizar inferencias válidas.

En síntesis, de acuerdo a estos autores, la lógica natural posee algún tipo de reglas que expresan cierto tipo de normatividad funcional, ya que si no la tuvieran no podría explicarse cómo un agente racional cualquiera, dadas las premisas A = B y B = C, infiere "con necesidad" la conclusión A = C.

Cualesquiera sean las propiedades elegidas para caracterizar la lógica natural, entendemos que las siguientes propiedades son las esenciales:

---

[3] Todos estos trabajos se encuentran en la obra *Reasoning, Necessity and Logic*, Eds. Willis F. Overton. Lawrence Erlbaum Associates, New Jersey, London, 1990.

1) La lógica natural no instrumenta, desde la sintaxis, cadenas de inferencias tan complejas como la que se dan en los sistemas de lógica formal y, contrariamente, los argumentos o inferencias de la lógica natural son más dificultosos de analizar desde una perspectiva semántica.
2) Los argumentos o razonamientos de la lógica natural tampoco siguen paso a paso una inferencia formal, sino que en ellos se suele pegar saltos, en cuya base están o bien la falta de información o la existencia de presuposiciones de la más diversa índole no explicitadas. Y
3) Las inferencias o razonamientos de la lógica natural permanecen generalmente ligados, por un lado, a la verdad o falsedad de los enunciados mismos o a las creencias que el hablante tiene acerca de ellos y, por el otro, al significado intensional de los enunciados involucrados y es en este sentido que se dice que los argumentos de la lógica natural son contexto- dependientes.

Luego de un breve comentario histórico que creemos ilustrativo, en la próxima sección nos ocuparemos de reseñar algunas propuestas que la lógica ha elaborado desde distintos enfoques a fin de adecuar más sus inferencias válidas a los razonamientos de sentido común.

II

Es sabido que la concepción deductiva de la lógica, en su versión aristotélica, perduró hasta el Renacimiento y, recién en 1555, tuvo su principal detractor en la *Dialectiqu*e de Petrus Ramus. En 1662, apareció la obra de A. Arnauld y P. Nicole, titulada *Logique ou l'art de penser*, conocida luego como *Lógica* de *Port Royal* con la cual se inicia un enfoque psicológico de la lógica cuya vigencia continua hasta 1870, momento en que la psicología se convierte en ciencia experimental y en el cual se critica la lógica aristotélica no por su objeto de estudio sino por la forma en que se lo concibe. Con la única excepción de Leibniz, quién, con sus ideas acerca de la posibilidad de existencia de una *característica universalis*, de un *ars combinatoria* y de un *calculus ratiocinator*, profundizó el carácter de cálculo formal de la lógica que aún se mantiene en nuestros días, para los lógicos de esa época, los conceptos, los juicios y los razonamientos seguirían siendo objeto de estudio de la lógica, pero ahora en tanto hechos mentales, sobre los cuales debería "reglar" la lógica, con el fin de diferenciar los pensamientos correctos de los incorrectos. Sin embargo, recién a principios del siglo XIX, con las ideas de Wundt, Sigwart y Lipps, se plantea con claridad el problema de la existencia de un pensamiento lógico natural y su normatividad, de la cual la lógica en tanto ciencia debería dar cuenta. De esta forma, la lógica "naturalizó" su objeto, en el sentido de que se convirtió en determinar cuáles son las normas y métodos correctos del pensar natural, dentro del cual queda incluido el pensar de sentido común. En síntesis, leído desde nuestra óptica, el conjunto de las operaciones de la mente constituiría para los psicologistas algo similar a lo que ahora llamaríamos lógica natural y la función de la lógica formal sería reglar sobre las

inferencias de la primera a los efectos de determinar su corrección. Sin embargo, señalamos que, pese a su naturalización, la lógica aristotélica siguió siendo el único formalismo vigente para determinar la corrección de las inferencias de tal lógica natural.

Recién en la segunda mitad del siglo XIX, con los aportes de Bolzano, Peirce, Boole y fundamentalmente Frege, la lógica retoma la línea de pensamiento que había iniciado Leibniz y con él se instala el Principio de Extensionalidad del significado que, si bien fue pensado por Frege para la matemática, se extendió posteriormente al análisis del lenguaje natural en las obras de Wittgenstein, Russell y Carnap. Dos son los aspectos fundamentales que queremos destacar en este punto: en primer lugar, la distinción entre forma gramatical y forma lógica y, en segundo lugar la idea de Russell que compartimos plenamente de que la forma gramatical no se corresponde siempre con la forma lógica, y que la primera puede resultar muchas veces engañosa. Desarrollos posteriores como los de Reyle y Strawson, contribuyeron a que J. L. Austin concluyera que el lenguaje natural no tiene una lógica precisa y que, por ello, no es posible analizarlo mediante reglas lógicas.

Nosotros no compartimos esta última posición y somos de la idea de que los distintos lenguajes naturales permiten construcciones que solamente son vehículos o medios que permiten expresar bajo formas lingüísticas distintas una misma proposición, cuya forma lógica será común a cada una de las primeras, lo cual conduce directamente a aceptar que detrás de los lenguajes hay una lógica natural cuya normatividad debería poder expresarse mediante un conjunto de "regularidades" lógicas, cuya naturaleza correspondería ser esclarecida por las ciencias cognitivas. Por otra parte, también coincidimos en lo ya reconocido por muchos filósofos de la lógica que la noción de validez intuitiva de un argumento no se corresponde necesariamente con su validez formal, ya que hay razonamientos que se consideran naturalmente válidos y sin embargo no son formalmente válidos. Ya lo hizo notar Carnap cuando mostró que el argumento "Juan es soltero, luego, Juan no es casado", si bien es intuitivamente válido no será lógicamente válido hasta que no se le agregue el postulado de significación que afirme "Todo hombre soltero es no-casado", dando de este modo uno de los primeros pasos de acercamiento entre lógica y pragmática.

Diversas lógicas, todas ellas deductivas y por lo tanto, monótonas, han surgido con la intención de adecuar las reglas de inferencia de la lógica clásica a las inferencias de la lógica natural. Una de las principales propuestas estrictamente formales provino de A. Ross Anderson y N. Belnap (A&B) en su obra de 1975 *The Logic of Relevance and Necessity* en la cual proponen sustituir la lógica clásica por una lógica de la relevancia (R), basándose en una fuerte crítica al condicional material clásico por la falta de relevancia significativa entre el antecedente y consecuente en las leyes y reglas de LC.

En primer lugar hacemos notar que la implicación relevante propuesta por A&B, tiene como objetivo reflejar en el lenguaje objeto de R la relación de deducibilidad relevante postulada en el metalenguaje. Se argumenta en contra de la adecuación de la noción de deducibilidad clásica porque ésta sólo involucra al concepto de necesidad en la derivación de la conclusión a partir de las premisas y bo toma en cuenta la relevancia entre premisas y conclusión. Esto implica aceptar como válidas inferencias no intuitivas, tales como la llamada *Falacia de la Relevancia* $A \vdash B \rightarrow A$ y

la regla del Pseudo Scotto (o E*x Contradictione Quodlibet*) $(A\wedge\neg A) \vdash B$. La exigencia de relevancia "significativa" es una consecuencia de la necesidad de expresar que en la deducción de una fórmula B a partir de una premisa A es relevante si y sólo si la fórmula A es usada en la deducción de B. En el lenguaje esta exigencia se traduce en el llamado *Principio de la Relevancia*, llamado también a veces *Principio de la comunidad de variables*, una de cuyas formulaciones dice que *Si A es un teorema de R, entonces toda variable que ocurre en A, ocurre al menos una vez en su parte antecedente y al menos una vez en su parte consecuente* (pp.34).

Sin embargo, esta propiedad es sólo condición necesaria pero no suficiente para determinar si una fórmula se sigue con relevancia significativa del antecedente, ya que existen fórmulas, como por ejemplo, $A\rightarrow(B\rightarrow(A\wedge B))$ y $((A\vee B) \wedge \neg A)\rightarrow B$ que comparten variables pero no son teorema de R.

Si bien se debe conceder que la exigencia de relevancia significativa es una propiedad importante en las inferencias de la lógica natural, esto se logra a costa de perder inferencias altamente intuitivas para la lógica natural como por ejemplo el Silogismo Disyuntivo, ya que su prueba no resulta relevante. Wójcicki, 1984,pp.43) muestra también que R no es una lógica bien determinada ya que admite considerar como inferencias válidas a ciertas deducciones no preservan la propiedad de compartir variables, como es el caso de la regla de R llamada *Cancelación de Conjunción válida* i.e., $A, (A\wedge B)\rightarrow C \vdash B\rightarrow C$, ya que, si en ella se sustituye la fórmula C por la fórmula $A\rightarrow A$, se obtiene como conclusión la fórmula no relevante $B\rightarrow(A\rightarrow A)$.

El intento de A&B para dar cuenta de ciertas relaciones significativas entre antecedente y consecuente no es el primero en la historia de la lógica. En efecto, W.Parry en el año 1933 había intentado formalizar la noción de implicación analítica, con el propósito de evitar las paradojas del condicional estricto y dar cuenta de la idea intuitiva de Kant de que un enunciado es analítico, o sea, que el enunciado A implica analíticamente al enunciado B si y sólo si el contenido significativo de A contiene al contenido significativo de B. Similarmente a lo que hemos mostrado sucede en el sistema R de A&B, en el sistema de Parry resultan inferencias inválidas las paradojas de la implicación estricta y *Ex Contradictione quodlibet.* Sin embargo, ahora resulta inválida la regla de Adición y válido el Silogismo Disyuntivo. Paradójicamente, pese a la intención del autor, resultan también fórmulas válidas, implicaciones muy alejadas de la idea de implicación analítica que se pretende formalizar y además, bastante contra intuitivas, como, por ejemplo las fórmulas, llamadas *Paradojas de la implicación analítica:* $\vdash(B\wedge\neg B)\wedge C\rightarrow\neg C$ y $\vdash(B\wedge\neg B)\wedge C\rightarrow(C\wedge\neg C)$

Otra de las limitaciones más conocidas de la lógica clásica y que se fundan también en el condicional material consiste en la imposibilidad de tratar en LC la inferencia contrafáctica. Es un hecho que en la argumentación de sentido común, el hombre efectúa inferencias a partir de enunciados contrafácticos (mal llamados "subjuntivos") y, sin embargo, la lógica clásica no puede dar cuenta de ellos, porque, traducidos al condicional material clásico, todos los enunciados contrafácticos resultan verdaderos.

Las respuestas que se dieron desde el campo de la lógica a esta problemática, estuvieron marcadas por el famoso Test de Ramsey, el cual sitúa el análisis de los condicionales en el plano pragmático del lenguaje natural. En efecto, en una nota a

pié de la página 247 del artículo de Ramsey, *General Proposition and Causality*, de 1931, éste afirma:
(a) (...) *Si dos personas están discutiendo acerca de "Si p entonces q" y ambas están en duda frente a "p", entonces están añadiendo hipotéticamente "p" a su conjunto de conocimiento y argumentando sobre esa base acerca de "q"; de tal forma que, en un sentido, "Si p, q" y "Si p,¬ q, son contradictorios. Podemos decir que ellas están ajustando sus grados de creencia en q dado p. Si p resulta ser falso, estos grados de creencia se vuelven nulos.*

Los primeros trabajos que trataron de dar cuenta de estos condicionales fueron los de N.Goodman en 1947 y R .M. Chisholm, en 1963. Sus teorías se agrupan bajo el nombre de *Teoría de la cosostenibilidad o Teoría Metalingüística*, se basaron en la siguiente idea: un condicional contrafáctico A>B es verdadero si y solo si A conjuntamente con un determinado conjunto Γ de leyes y proposiciones verdaderas implican (clásicamente) a B, o sea:

$$A>B \text{ es Verdadro sii } A \wedge \Gamma \rightarrow B$$

Como se ve claramente, la estrategia consistió en agregar un conjunto Γ de enunciados que en conjunción con el antecedente A fuera condición suficiente para B. El principal problema de este enfoque fue obviamente cómo determinar el conjunto Γ de enunciados de tal forma que ningún enunciado de Γ implicara ¬B, es decir, que el conjunto Γ no hubiera ningún enunciado que derrotara al consecuente.. Dado el tradicional ejemplo contrafáctico:

Si el fósforo hubiera sido raspado, se habría encendido.

Es obvio que en el conjunto de los enunciados verdaderos Γ que deberían agregarse al antecedente a los efectos de obtenerse una condición suficiente para el consecuente, deberían encontrarse todos los enunciados relevantes cuya conjunción fuera condición suficiente para el consecuente. En particular el enunciado *el fósforo hubiera estado seco* debería pertenecer al conjunto Γ y no podría encontrarse en Γ el enunciado que afirmara que el fósforo estaba húmedo.

Las soluciones puramente lógicas fracasaron pues se mostró que los criterios propuestos desde la lógica como, por ejemplo, que los enunciados que componen Γ debían ser lógicamente independientes o que ninguno de ellos debía implicar la negación del consecuente, resultaron insuficientes y convertía inevitablemente la argumentación en circular. En síntesis, la selección de cuál información relevante se debe agregar para completar el antecedente a fin de hacer que éste sea una condición suficiente, no es determinable por vía puramente lógica, sino que implica agregar información no obtenible por métodos lógicos sino pragmáticos.

Las semánticas de mundos posibles construidas por Kripke para las lógicas modales posibilitaron análisis nuevos para los condicionales contrafácticos posibilitando la construcción de ricos sistemas lógicos que son considerados extensiones de la lógica clásica. Estos se construyeron, o bien introduciendo en el lenguaje objeto una función *f* de selección (R. Stalnaker, 1975 "A Theory of Conditionals") la cual para cada enunciado condicional A>B selecciona el mundo o el conjunto de mundos en el cual el antecedente A es verdadero, o introduciendo en el metalenguaje semántico (D.Lewis, 1973, *Counterfactuals*,) un sistema de esferas de

mundos posibles accesibles y ordenado de acuerdo a una relación de similaridad comparativa respecto del mundo actual. Es evidente que en ambos enfoques se reproduce un problema no solucionable por métodos lógicos únicamente, ya que no es una cuestión a resolver por la lógica qué mundo o qué conjunto de mundos accesibles elige la función *f*, o bien cuáles son los mundos posibles más similares al actual a los cuales hay que acudir para determinar el valor de verdad de un enunciado condicional contrafáctico. Se hace evidente que el carácter pragmático de los requisitos mencionados convierte a estas lógicas en contexto-dependientes.

Pese a la diversidad de sistemas condicionales concebidos como extensiones de la lógica clásica (modal o no), en todos ellos no valen tres reglas de inferencia clásicas, a saber: El Refuerzo del Antecedente (RA), la Contraposición (CN) y la Transitividad (TR), o sea:

(RA): $A \rightarrow B \nvdash (A \wedge C) \rightarrow B$

(CN): $A \rightarrow \neg B \nvdash B \rightarrow \neg A$

(TR): $A \rightarrow B, B \rightarrow C \nvdash A \rightarrow C$

Interesa a nuestros propósitos RA, ya que ella refleja en el lenguaje objeto la propiedad de Monotonía (deducibilidad) de la noción de consecuencia lógica clásica. Del ejemplo de enunciado contrafáctioco ya dado *Si el fósforo hubiera sido raspado, se habría encendido*, no se sigue preservando la verdad *Si el fósforo hubiera sido raspado y hubiera estado húmedo se habría encendido*, enunciado sin duda alguno verdadero para la lógica natural de cualquier hablante racional.

De esta forma, a la crítica proveniente de la falta de relevancia significativa de la lógica para dar cuenta de las inferencias de la lógica natural, se agregan ahora nada más y nada menos que la crítica a su carácter deductivo. En otras palabras, si quisiéramos encontrar un sistema de reglas básicas que conformaran el núcleo central de la lógica natural, éste no parecería ser un sistema deductivo. Y es aquí donde precisamente se junta la problemática de la lógica clásica y la de la Inteligencia Artificial, la cual pasaremos ahora a reseñar brevemente.

III

El mismo año en el que aparece el libro de Anderson y Belnapen en el cual se criticaba la incapacidad de LC para dar cuenta de la relevancia deductiva, i.e.,1975, se publica el libro de M. Minsky *A Framework for Reoresenting Knowledge*, en el cual expone los aspectos de LC que considera inadecuados para representar los razonamientos de sentido común, entre los que nos interesa destacar los siguientes por su carácter esencialmente lógico: (i) la ausencia de relevancia, (ii) la propiedad de monotonía y (iii) la exigencia de consistencia en LC expresada por la validez de la regla ECQ. Respecto de las dos primeras ya hemos reseñado las propuestas dadas por la lógica que consideramos más importantes. Ahora deseamos agregar que el tercer problema también fue contemplado desde la lógica formal. En efecto, ya en 1912, el Principio de Duns Scoto fue criticado por el lógico aristotélico N. A. Vasil'ev. Más aún, los primeros sistemas formales, de ahora en más llamados *paraconsistentes* aparecieron una vez terminada la segunda guerra mundial, en lugares muy distintos y

en forma independiente unos de otros. El primero fue creado por el lógico polaco S, Jáskowski, en 1949 y los restantes se deben a F. G. Asenjo, (Argentina1954), N.C.A. da Costa (Brasil, 1958) y T. J. Smiley (Reino Unido,1959). El término "paraconsistente" fue propuesto por el peruano F. Miró Quesada ya mencionado en el 3er. Simposio Latinoamericano sobre lógica matemática, celebrado en el año 1976. Sin embargo, sólo en Brasil, con la escuela de C. A. Newton da Costa y en Australia con G. Priest puede afirmarse que se ha formado una tradición en este tipo de lógica. Todos los ellos justifican la necesidad de construir sistemas de lógica paraconsistente, que admitan proposiciones contradictorias pero que al mismo tiempo no permitan que en ellos cualquier afirmación sea demostrable, i.e., que sean trivialmente inconsistentes. En general las razones dadas para postular estas lógicas son: 1) la inconsistencia es un fenómeno natural del mundo que se manifiestan en conjuntos de informaciones, conjuntos de leyes u obligaciones morales, de creencias e incluso en teorías científicas, como es el caso de la física cuántica; 2**)** la existencia en el lenguaje natural de paradojas semánticas y de predicados vagos, puesto que para estos casos es plausible admitir como consecuencia enunciado contradictorio, como por ejemplo *Pedro es pelado y no pelado*; y 3) razones propiamente filosóficas, tales como las insinuadas por Wittgenstein en sus escritos sobre matemática cuando afirma que las contradicciones no son tan destructivas como los formalismos clásicos creen. Sin embargo, pareciera que el rechazo o la admisibilidad de contradicciones depende, en última instancia de lo que habrá de entenderse por la negación. Desde la lógica, se han caracterizado siete formas de negación, pero no se ha podido responder a la pregunta de qué es la negación. Al respecto, citaremos un ilustrativo párrafo de M. Dunn:

"Qué es la verdad?", preguntó Pontius Pilates. "Qué es la negación? preguntan Dov Gabbay y Heinrich Wansing. Pilates nunca obtuvo una respuesta, y yo no puedo responder la pregunta de Dov and Heinrich, a menos que una cantidad de *respuestas pueda ser considerada como la respuesta. En su lugar yo mostraré cómo a partir de variadas propiedades estructurales de la negación es posible obtener varias perspectivas modelo-teoréticas.* [4]

Si esta es la situación de la negación en la lógica, qué podríamos esperar del comportamiento de la negación en el lenguaje natural. La obra más completa sobre la negación que se conoce es la de Laurence R. Horn (1989) *A Natural History of Negation.* En ella se muestra que, a pesar de su temprana aparición en el desarrollo de la inteligencia y de la simplicidad de su caracterización en la LC y en otros sistemas formales, la situación que se presenta en el uso de la negación en los lenguajes naturales es absolutamente la inversa y no se ha llegado aún a ningún esclarecimiento preciso de sus distintos usos en los diferentes tipos de discurso ni desde la lingüística, ni desde la psicolingüística. Al decir de Laurence Horn, la negación es para los lingüistas y filósofos del lenguaje como un fruto de Tantalus: *ondula seductoramente,*

- [4] "AComparative study of Various Model-theoreticTreatements of Negation: A History of Formal Negation", en *What is Negation,* D.M.Gabbay y H. Wansing eds, Kluwer, Academic Publishers, pp.48.

tentadoramente palpable, presta a ser alcanzada solamente cuando está pronta a escapar una vez más [5].

En la actualidad existen diversos sistemas de lógicas paraconsistentes para ser aplicadas en I.A. Sin embargo, estos sistemas no agregan claridad alguna al comportamiento de la negación natural, habida cuenta que tampoco está claro el significado de la negación lógica en los sistemas de lógica paraconsistente.

Retomemos ahora la cuestión de cómo continúa el análisis de las inferencias derrotables en el campo de la I.A. en lo que respecta a la representación del llamado *conocimiento de sentido común.*. Es sabido que la idea de usar en I.A. la lógica como una representación subyacente de los razonamientos de sentido común se inició con el trabajo de 1959 de J. Mc Carthy, titulado precisamente *Programs with Common Sense.* Es conocida la cantidad de formalismos construidos desde I.A. para dar cuenta de los razonamientos con premisas plausibles o con excepciones o con información incompleta. Creemos posible afirmar que las contribuciones que más contribuyeron a la sistematización lógica de la inferencia no monótona son: la intepretación procedural de las cláusulas Horn (Kowalski, 1974), el desarrollo del lenguaje de programación lógica PROLOG (Colmenauer y Roussel, 1972), la Hipótesis del Mundo cerrado (R. Reiter,1978); Circunscripción (McCarthy,1980) y *A Logic for Default Reasoning* (R. Reiter,1980) ya que todos ellos constituyen formalismos no monótonos, o sea "derrotables". También se coincide que fue Dov M. Gabbay quien, en 1985, estableció por primera vez las propiedades generales del razonamiento de sentido común o, si se prefiere, los razonamientos de la lógica natural, en su trabajo *Theoretical foundations for non-monotonic reasoning in expert systems.* Años después, en 1990, aparece el trabajo de de S. Krauss. D. Lehmann y M. Magidor (KLM), titulado *Nonmonotonic Reasoning, Preferential Models and Cumulative Logics*, y, en 1994, en el cual se da una primera versión de la noción de consecuencia involucrada en ese tipo de consecuencia. En 1994, David Makinson presenta la más acabada versión de los trabajos sobre este tema en *General Patterns in Nonmonotonic Reasoning* y en su obra reciente de 2005, *Bridges from Classical to Nonmonotonic Logic* en el cual formula las diversas formas de consecuencia no monótona de los formalismos tradicionales, establece las relaciones entre ellas y analiza las conexiones entre la no monotonía y la probabilidad lógica. A diferencia de los sistemas subclásicos, los sistemas de lógica no monótonos son supraclásicos o sea que se construyen a partir de la lógica clásica añadiendo un signo para la relación de consecuencia no monótona |~, más un conjunto de reglas de inferencia. Hoy en general se acuerda en considerar básicas las siguientes reglas:

- [5] Horn, Laurence,2001, *A Natural History of Negation*, The David Hume Series. CSLI Publications.

$A \mid\sim A$ *Reflexividad* (R)

$$\frac{A \mid\sim B \quad A \wedge B \mid\sim C}{A \mid\sim C}$$ *Corte o Transitividad Cumulativa* (TC)

$$\frac{A \mid\sim B \quad A \mid\sim C}{A \wedge B \mid\sim C}$$ *Monotonía cautelosa* o *cumulativa* (MC)

Es posible también agregar otras reglas, como por ejemplo, la llamada *Monotonía Racional*, de forma tal que distintos conjuntos de reglas conformarán distintas nociones de consecuencia lógica no-monótona.

En líneas generales, las diferentes semánticas propuestas para estos sistemas no monótonos tratan de rescatar la idea de McCarthy de dominios mínimos o extensiones mínimas de predicados. De ahí que, en lugar de todos los modelos, tal como se hace en los sistemas modales clásicos, se tome en cuenta solamente un tipo particular de modelos, según el tipo de *minimalidad* semántica que se elija. Por ello, estos modelos son llamados *modelos mínimos* o, si se tiene en cuenta el orden de preferencia entre estados, *modelos preferenciales* o, si se quiere, *modelos normales*, lo cual pone de manifiesto el carácter contexto dependiente de los sistemas no monótonos análogamente a los sistemas de lógica condicional citados.

Preguntémonos ahora si las nociones de consecuencia lógica no monótonas ayudan a elucidar las reglas que están involucradas en los razonamientos de la lógica natural, para lo cual nos es suficiente analizar el comportamiento de TC y MR.

La transitividad cumulativa (TC) y Monotonía cautelosa (MC) parecen comportarse efectivamente como lo hace un agente racional, en el sentido de que, aún cuando sus premisas sean inferencias plausibles, en el razonar común el agente conserva las conclusiones obtenidas en sus inferencias anteriores. Si se interpreta A como: *Juan obtiene una beca para estudiar en España*, B como *Juan se radicará en España* y C como *Juan formará una familia en España*, se obtienen efectivamente inferencias naturales que satisfacen ambas reglas.

Sin embargo, David. Makinson, en *Bridges from Classical to Nonmonotoic Logic*, (2005)[6], no todos los formalismos no monótonos de la IA, o sea, no todos los argumentos de sentido común satisfacen las mismas reglas, originándose diferentes consecuencias no monótonas. Dada la complejidad que involucraría la presentación formal de estos resultados, daremos solamente dos ejemplos. Si a un conjunto de creencias (o enunciados) K se le agrega una información A que es inconsistente con K entonces $C_K(A)$ es no-monótona, depende sintácticamente de los elementos de K y satisface Monotonía Cautelosa(MC), tal como sucede en la Hipótesis del Mundo Cerrado de Reiter. Pero, si se toma la lógica por defecto de Reiter (en su

[6] Text in Computing, vol 5, 2005,King's College, London.

interpretación escéptica) su noción de consecuencia $|\sim_R$ es no monótona pero no satisface Monotonía cautelosa (MC).

Finalmente, la relación de consecuencia no monótona tiene dos propiedades negativas a señalar que comprometen su carácter de ser consideradas una lógica: 1) generalmente no satisface compacidad y 2) no es cerrada bajo la regla de Sustitución Uniforme. Si bien estas características son aceptables, si se piensa que razonar no monótonamente es obtener conclusiones que son modificables cuando se agrega nueva información, fundamentalmente la no satisfacibilidad de Sustitución Uniforme atenta directamente contra la posibilidad de considerarla una noción de consecuencia lógica formal. Tal vez por ello Makinson considera que tomar Sustitución Uniforme como criterio de logicidad es un "un hábito a suspender" en los estudiosos de las inferencias del sentido común.

Para finalizar, partiendo de que aceptamos que la lógica clásica no resulta totalmente adecuada para formalizar los razonamientos de la lógica natural en todo dominio de discurso, que una correcta aplicación de ella a dominios específicos requiere la introducción de restricciones pragmáticas de la más diversa índole, que los sistemas de lógica subclásicos complican sintácticamente las reglas inferenciales y que la teoría de la consecuencia no monótona, pese a ser adecuada en cuanto a la descripción de la forma de los razonamientos de sentido común, no proporciona ningún criterio unívoco para determinar la corrección de los mismos, somos de la opinión que la manera más optima, al menos por ahora, de tratar los razonamientos de sentido común, es la propuesta por los expertos en IA, en particular cuando representan las inferencias mediante reglas de la lógica clásica y agregan mecanismos de control pragmático, ya sea como supuestos o determinando sus dominios de aplicación, cerrando de esta forma el círculo con la sugerencia de Carnap mencionada al comienzo de nuestro trabajo.

Usando palabras de Donald Gillies en su libro *Intelligence and Scientific Method* (1996,pp.85)

Cuando empleamos Lógica partimos de un conjunto de supuestos a partir de los cuales se derivan ciertas conclusiones. Para llevar a cabo tales derivaciones es necesario contar con un conjunto de reglas de inferencia (el componente Inferencial). En adición a estas reglas de inferencia, necesitamos generalmente alguna guía práctica para decidir qué suposiciones elegir y qué reglas de inferencia utilizar. Así, el componente Control ayuda en la construcción de la derivación para alcanzar una conclusión válida.

Aspectos Filosóficos de la Inteligencia y la Computación

E. Alonso

Universidad Autónoma de Madrid

enrique.alonso@uam.es

1 ¿En qué consiste un estudio *multidisciplinar*?

Confieso que la tentación es grande y superior a lo que soy capaz de resistir. Por tanto diré algo acerca de qué pienso que es un estudio multidisciplinar, o mejor dicho, qué me parece que no debe ser en modo alguno.

Hace ya algún tiempo que términos como *multidisciplinar* o *interdisciplinar* irrumpieron con fuerza entre nosotros llegándose a convertir en adendas convenientes a todo proyecto que aspirase a encontrar su lugar bajo el sol. Para que una iniciativa pudiera aspirar a gozar del favor de las instituciones científicas o académicas nacionales era conveniente, si no necesario, incluir tales términos en lugar bien visible, dedicando unas pocas palabras a mostrar con convencimiento el carácter propiamente multidisciplinar del proyecto en cuestión.

Bajo esta tendencia se oculta un hecho manifiesto que no se resuelve, o eso me temo, con maniobras de mercadotecnia como la que supone la mera adición de una palabra más o menos simpática. El hecho al que me refiero es la escasa tradición existente en nuestro entorno a la hora de discutir y compartir problemas. Pese al indudable avance experimentado por la investigación y la práctica científica en nuestro país durante las últimas décadas, aún no hemos aprendido a discutir entre nosotros sobre aquello que hacemos. Por tanto, un estudio multidisciplinar no es uno que lo sea sólo por su tema, sino aquel en el que se ha dejado hablar al *otro* y sobre todo se le ha *escuchado*.

Pese a la intención manifiesta, los estudios multidisciplinares no consiguen transmitir la impresión de admitir en sus debates a profesionales procedentes de otros ámbitos y materias. Se supone que un estudio multidisciplinar se aplica a sí mismo esa etiqueta para indicar claramente su voluntad de trascender el estricto ámbito de lo *disciplinar*. Pero lo cierto es que veces lo que se logra es provocar la impresión de que para saber de *eso* lo que hace falta es saber de *todo*. Un ejemplo lo aclara. La Historia de la Ciencia y la Filosofía de la Ciencia son materias que forman parte de la formación curricular básica del filósofo. No se trata de materias dirigidas a científicos que deseen plantearse su disciplina desde otro punto de vista. No se supone una formación científica completa ni se aspira a que el estudiante la obtenga. No obstante, es frecuente, casi recurrente, tener que explicar que tales materias no están reservadas a genios universales capaces de abarcar dos o tres ámbitos del saber con parecida competencia. Nadie se descarta de alcanzar la máxima relevancia en este tipo de materias, las típicamente multidisciplinares, por ignorar parte de aquello que se ve

A. Fernández-Caballero & S. Miguel (Eds.): 50 Años de la Inteligencia Artificial, pp. 25-41, 2006.

implicado en su mismo nombre. Este no puede ser el modo de entender el carácter multidisciplinar de un estudio o problema.

Las reuniones en que varios profesionales, procedentes de ámbitos distintos, fijan objetivos dentro de un marco multidisciplinar se parecen en ocasiones a los planes de invasión de un territorio. El ejército de tierra se ocupará de estos y aquellos objetivos, la Armada de las costas y suministros marítimos, la aviación del control de aeropuertos y vías de comunicación. Se actúa repartiendo, según ámbitos de competencia, aquellas tareas que se ven implicadas en el total del proyecto. En nuestro caso, si el filósofo o el lógico tienen una idea que atañe al ámbito de la computación su deber, como genuino científico multidisciplinar, es poner su proyecto en conocimiento de la comunidad de técnicos informáticos encargados de forma natural de su implementación. Si el ingeniero tiene dudas filosóficas lo suyo es acudir al filósofo, quien en una sesión de diván procurará que éstas queden resueltas orientando en todo caso sus pesquisas y aportando las lecturas convenientes para ello. Pero, también es posible que el filósofo sepa programar y que el ingeniero sepa buscar en una biblioteca y sea capaz de leer y entender aquello que ha encontrado.

Hay veces en las que el reparto de tareas representa una buena medida a la hora de abordar un problema. De hecho, las más. Hay situaciones en las que incluso no se puede proceder de otro modo, pero plantearse este logro como punto de partida se parece demasiado al reparto de distritos y regiones que podemos recordar de clásicos del cine dedicados al mundo del *hampa*.

Para finalizar me gustaría recordar que lo que los científicos, filósofos y otros profesionales del saber tratan de resolver son problemas. Los problemas no son, aunque nos hayamos acostumbrado a verlo así, propiedades de las disciplinas que intentan repartirse el mapa del conocimiento. Es cierto que cuando encaramos un problema a menudo ponemos por delante de otras consideraciones la correcta identificación de los derechos de propiedad que cada disciplina pueda llegar a reclamar. Sin embargo, esta no me parece la actitud a seguir. Los problemas, si son de alguien, es de aquellas personas que los plantean y por supuesto de las que los resuelven. Si hay alguna recomendación que yo mismo me hago a la hora de abordar una determinada cuestión es olvidar el marco disciplinar y dejar que sea el problema el que hable mostrando aquello de que está hecho. Esta es, desde mi punto de vista, la única forma genuina y aceptable de entender el carácter multidisciplinar, no ya de este o aquel asunto, sino del conocimiento como un todo. La actitud que me impongo es esta: el conocimiento es, mientras no se diga otra cosa, una empresa multidisciplinar capaz de sacarnos de nuestras casillas –*disciplinas*- siempre que sea preciso, siempre que a la naturaleza del problema así se le antoje. Por tanto, más nos vale estar en continua disposición de mudanza, de aprender nuevos métodos y recursos para aquello que se precise. Y que nadie crea que recomiendo algo así como la *formación continua*, que más parece destinada a convertirnos en *multititulados* que en personas de mente abierta. Porque, ¿qué cantidad de conocimiento del que aprendemos a lo largo del duro estudio de una Licenciatura aplicamos después en aquello que hacemos y realmente nos importa?

2 Lo que importa de la Lógica y la Filosofía.

Para que no parezca otra cosa diré ahora que el contexto en que esta iniciativa se ha desarrollado me parece bastante a salvo de las malas interpretaciones de la multidisciplinariedad que acabo de relatar. Intentaré hacer yo también lo propio.

A la hora de preparar un texto como este se plantean muchas opciones. ¿Cuál de ellas es la que mejor puede satisfacer los objetivos que nos hemos venido a plantear aquí? Supongo que aquella que más relevante pueda resultar a las personas que menos cerca se encuentren de nuestras preocupaciones disciplinares. No se trata de contar lo último que hemos hecho, sino de explicar a aquellos que hacen otras cosas qué les puede interesar de lo que uno hace. ¿Qué le puede importar a una persona interesada en el proyecto general de la I.A. lo que un lógico o un filósofo haya hecho o dicho a ese respecto? Esta es la cuestión que intentaré explicar aquí. Y que quede claro que no pretendo regañar a nadie por no saber de antemano qué le puede interesar de estos asuntos, ni me iré a disgusto tampoco si al final aún hay personas que no creen que merezca la pena dedicarle más tiempo al tipo de problemas que a mi me importan. Se trata de mostrar caminos, no de obligar a nadie a transitar por ellos.

Empezaré por lo que me es más próximo, la Lógica. Para entender lo que la Lógica tiene que ver con el problema general de la I.A. quizá convenga recordar algo de la historia concreta del nacimiento de este programa de investigación. Hay avances técnicos que a menudo preceden a cualquier teoría capaz de explicar su funcionamiento, éxito o propiedades más elementales. Gran parte de la mecánica se entiende sólo como un intento de explicar el rendimiento de artefactos que en ocasiones existen prácticamente desde el inicio de la civilización. En tiempos más recientes contamos con el ejemplo de la termodinámica cuyo fin inicial es explicar el rendimiento térmico de las máquinas que impulsaron la revolución industrial. En todos estos casos podemos hablar, con todos los matices que se quieran introducir, de un modelo en el que *el artefacto precede a la teoría*. ¿Es este nuestro modelo? No, ciertamente. En el caso de la I.A. ésta nace sólo después de que la Teoría de la Computación hubiera puesto sobre la mesa ciertas posibilidades teóricas explicadas de un modo que probablemente no tiene precedentes en el tiempo. Los principales resultados en computación teórica datan de la segunda mitad de la década de 1930 mientras que, si aceptamos la Conferencia de Dartmouth como hito inaugural del proyecto de la I.A., nos vemos impulsados hacia mediados de la década de 1950. De hecho, los artículos seminales en Computación, firmados el primero por A. Church y el segundo por A. Turing datan ambos de 1936, 20 años exactos antes de la celebración de la Conferencia de Dartmouth.

Los conceptos que abren paso al desarrollo de la I.A. como un proyecto serio y guían de hecho la construcción de los primeros ordenadores son, aunque ahora sólo lo mencione, el de *máquina de Turing*, entendido como modelo general y abstracto de lo que supone ejecutar un tarea de manera efectiva, y su compañero de viaje, el concepto de *máquina universal de Turing*. Hablaré de ello más adelante.

Estaríamos, pues ante un ejemplo en el que la *teoría precede al artefacto*, algo similar a lo que también en aquella época sucediera con la energía nuclear. La importancia que este hecho tiene aquí se debe a que muchas de las discusiones teóricas que se llevan a cabo en relación con la posibilidad misma de la I.A. giran

entorno a aspectos concretos del modelo teórico en que están basados los ordenadores, es decir, tratan en definitiva de máquinas de Turing. Y todo ello reposa a su vez en consideraciones típicas de la Lógica. La Teoría de la Computación nace de hecho a partir de las respuestas que autores como los que acabo de mencionar dan a una serie de problemas vigentes en la Lógica del primer tercio del siglo xx.

No quiero que lo dicho eclipse por completo la importancia que el tratamiento de estos asuntos ha tenido en la tradición occidental. El problema de si las capacidades humanas pueden ser reproducidas por otros medios es una constante en todas las épocas del desarrollo intelectual de occidente. Basta con repasar sus mitos –homúnculos y Golem, hombres de palo y bustos parlantes- y también algunos de sus logros manifiestos para entender lo profundamente arraigada que se encuentra en nuestra cultura la sospecha fundamental en torno a nuestra pretendida originalidad –la del ser humano-. No obstante, sí me gustaría decir algo en relación a aquellos ingenios que realmente llegaron a ejecutarse o al menos a proyectarse.

Las máquinas calculadoras de Leibniz o Pascal suelen citarse en ocasiones como posibles precursores del estudio de la automatización de nuestros procesos cognitivos. No obstante, el parecido con los recursos de derivados del nacimiento de la Teoría de la Computación es tan lejano que no merece la pena entretenerse en ello. Mucho más interesante es el ejemplo recurrente de la *Analytical Engine* diseñada –que no construida- por Babbage. En este caso existe una sorprendente similitud de diseño con nuestras máquinas actuales. Estoy dispuesto a aceptar incluso la existencia de una cierta anticipación de la idea de *sistema operativo* y con ello de *programación*, pero lo que no puedo conceder es una similitud esencial, no ocasional, con el modelo hoy vigente. Y ello se debe a que la teoría que soporta el diseño de la Analytical Engine es por entero distinta a la teoría formal y abstracta en que se apoya la Teoría de la Computación. La Analytical Engine, como cualesquiera otros mecanismos diseñados, fabricados o imaginados antes de 1936 sólo podía encontrar apoyo teórico en la mecánica racional newtoniana, o a lo sumo en el electromagnetismo clásico. Nada que ver con el banco de métodos, intuiciones y problemas con que ahora contamos.

Una vez aclarado esto, volvamos al punto. ¿Qué puede aportar la Lógica al debate en torno a la I.A.? A mi entender hay dos grandes ámbitos en los que al Lógica puede ser de alguna ayuda.

El primero se refiere al estudio y análisis de lo que voy a llamar los *argumentos mirabilis*. Estos argumentos son todos aquellos que intentan encontrar razones puramente conceptuales, que los filósofos solemos llamar *a priori*, para mostrar, o mejor *de*mostrar, la superioridad de la mente humana sobre cualquier ingenio mecánico y de ahí la imposibilidad de alcanzar su potencia expresiva por medios artificiales. Por extraño que parezca se trata de un fenómeno recurrente capaz e rearmarse cada cierto tiempo para volver al debate bajo alguna variante más o menos llamativa. La Lógica es extremadamente eficaz a la hora de desmontar este tipo de argumentos porque seguramente es ella misma la fuente de todos ellos.

El segundo ámbito en el que la Lógica puede ofrecer ventajas a aquellos que se interesen por sus métodos es el del estudio de posibles alternativas al modelo computacional clásico, lo que se viene llamando *computación no-clásica.* Debe tenerse en cuenta que toda diferencia real entre presuntos modelos computacionales alternativos y rivales se resuelve en última instancia demostrando la existencia de un procedimiento computable en un sentido pero no en el restante. Estas demostraciones

son, si algo lo es, demostraciones *lógicas*. Conocer la armazón lógica del modelo vigente permite o facilita al menos, el trazado de las condiciones de contorno de cualquier modelo que realmente aspire a *hacer* algo que el modelo clásico no pueda. No siempre es fácil entender en qué consisten esas condiciones o rasgos básicos.

Una vez aclarado el punto referente a lo que la Lógica puede hacer por una mejor comprensión de los problemas de la I.A. le toca el turno a la propia Filosofía. En este caso confieso que mis certezas son, si es que existen, mucho menores que en lo que atañe a la Lógica. El volumen de reflexión filosófica que de un modo u otro tiene aplicación al tipo de problemas que aquí nos importan es de tal magnitud que en ocasiones parece sobrepasar la del propio objeto al cual se refieren. ¿Cómo orientarse en este aparente caos? Admito que no lo se. Mi propuesta en este caso es una mera opción sin otras pretensiones. Me limitaré a relatar lo que ha sido mi propia experiencia analizada esta vez de un modo muy general.

Como en el caso anterior, seleccionaré dos ámbitos en los que la reflexión filosófica me parece relevante para nuestros objetivos. El primero se refiere al valor de lo que solemos denominar *experimentos mentales*. Un experimento mental es un tipo de argumento en el que se plantea una situación experimental ficticia cuyo desarrollo permite apoyar una cierta conclusión. En ocasiones estos experimentos pueden llevarse realmente a la práctica -aunque no se haga- mientras que en otras resulta simplemente imposible. Ejemplos famosos de este tipo de argumentos son el de la *piedra y el mástil*, en que se establece el principio de inercia, el de la *pluma y el bloque de hierro*, orientado a mostrar la independencia de la aceleración de un móvil en caída libre con respecto a su masa, y los que aquí nos interesan de forma directa: el *Test de Turing*, la *Habitación China*, *Aquiles y la Tortuga* o la *Lámpara de Thomson*. En todos estos casos hay una tesis cuya defensa se basa en el propio desarrollo conceptual del experimento. La importancia de este tipo de argumentos en situaciones en las que no contamos con la posibilidad de generar prácticamente las condiciones experimentales requeridas es fundamental para romper prejuicios o para iluminar nuevos aspectos de un problema. Pero también hay riesgos en ello, no quizá de sufrir un accidente al manipular indebidamente un material o herramienta, pero sí de quedarse atrapado en un sinfín de consideraciones crecientemente irrelevantes para lo que originalmente se pretendía establecer.

El otro frente en el que la Filosofía me resulta de especial ayuda es más difícil de catalogar. Hay ocasiones en las que es preciso y conveniente abrir los oídos a fenómenos e ideas que poco a nada tienen que ver con aquello que uno hace. No se trata de tener una especie de hobbie o segunda ocupación sino de tomar aire y referencias en dominios cuyas normas de conducta son por entero distintas a aquellas a las que uno está acostumbrado. ¿Cuántas veces hemos oído hablar de que tal o cual idea le sobrevino a su autor de forma accidental tras un viaje a tierras extrañas o tras la lectura de una obra de ficción o la contemplación de cierta obra de arte o, lo que es incluso más frecuente, tras un simple sueño? La filosofía es a veces extremadamente útil como fuente de intuiciones capaces de desbloquear situaciones de aparente impasse. En este sentido no difiere sustancialmente de lo que podría ser un mero repaso a la historia de las ideas, algo a lo que no suele prestarse la debida atención.

Entendida como museo de Historia Natural del Pensamiento la Filosofía contiene salas en las que merece la pena perder una tarde, o más de una. Destaco tan sólo tres. La primera está dedicada al fenómeno de la conciencia. Analizaré, mencionaré más

bien, algunos prejuicios relativos al modo en que este problema incide en el desarrollo de la I.A. El segundo problema que me interesa tiene que ver con lo que se ha denominado actitud *holista*. Veremos de qué modo el holismo ha ido ganando terreno en la interpretación del proyecto de la I.A. al punto de ser el fundamento de la filosofía de los sistemas multiagente en oposición a la I.A. formal o clásica. El tercer y último aspecto de mi visita parte del ámbito de la *paleoantropología* y se refiere al hecho básico de la *especiación*. La I.A. presenta en ocasiones una gran dependencia de modelos fuertemente antropocéntricos de la inteligencia. El estudio de algunos de los hechos básicos de la paleontología puede contribuir en algo a corregir esta tendencia aceptando interpretaciones mucho más amplias del fenómeno general de la inteligencia y sobre todo de la *existencia artificial*.

Estas son mis intenciones, empecemos pues con su desarrollo.

3 La increíble tradición de los *argumentos mirabilis*.

Nuestra tradición intelectual muestra lo que podría considerarse como una serie de actitudes fundamentales en constante proceso de debate y renovación. Si nos preguntamos, por ejemplo, acerca de la fuente principal del conocimiento nos veremos involucrados en un debate que antes o después acabará por reproducir dos de estas actitudes. Una verá en la experiencia el origen fundamental de todo conocimiento posible, la otra hará de la razón y su capacidad de análisis el origen básico de nuestras certezas. Como es obvio existen opciones intermedias muy elaboradas y seguramente mejor orientadas a la verdad que estas posiciones límite, pero no es eso lo que deseo discutir ahora. Si planteo este problema es porque tiene mucho que ver con las formas de criticar el proyecto de la I.A. ¿De qué maneras se puede atacar la posibilidad de fabricar inteligencia artificial? De las muchas posibles podemos extraer dos opciones a su vez extremas. La primera vía la ofrece el *escepticismo*. Los argumentos escépticos suelen ser eficaces en aquellas personas previamente dispuestas en contra de la I.A. pero en un cierto sentido son argumentos débiles. El escéptico intentará mostrar la implausibilidad de la I.A., su carácter contraintuivo, sus inconveniencias, y cualquier otra clase de reparo imaginable. Esto no significa que la actitud escéptica no sea eficaz, lo único que quiero mostrar al hablar así, es que el argumento del escéptico nunca es concluyente o probatorio. La segunda opción es por entero distinta ya que se orienta, precisamente, a buscar un argumento definitivo, necesario, concluyente contra la I.A. A menudo se usa el término *mentalista* para caracterizar a aquellos que apuestan por esta opción, pero lo cierto es que se puede ser mentalista en un sentido mucho más débil que el que ahora me interesa: se puede ser mentalista y aportar sólo argumentos escépticos. El mentalista extremo confía pues en la capacidad de la Razón para hallar un argumento necesario, objetivo y probatorio capaz de desbaratar los esmeros que preocupan a tantos. Para ello tiene que confiar fuertemente en la capacidad de la razón humana para extraer de las condiciones en que se establece un cierto problema datos relevantes para su solución. Me he permitido la licencia de referirme a este tipo de

argumentos como *argumentos mirabilis* por lo mucho de sorprendente que hay en ellos, pero tampoco pretendo ridiculizarlos sin más. En ocasiones son elaboraciones de una sutil belleza constituyendo piezas de estudio de considerable valor.

El hecho de que podamos hablar en plural de este tipo de argumentos se debe a la existencia de una larga tradición que se extiende tanto en el tiempo como en lo que se refiere al contenido. La variante más extrema y comentada de argumento mirabilis lo constituye el denominado *argumento ontológico* destinado a demostrar la existencia de Dios de forma racional. Existen variantes de todo tipo siendo las más famosas la de San Anselmo y la del propio Descartes. Resulta sorprendente, y no muy sabido, el hecho de que el propio Gödel desarrollara uno de estos argumentos haciendo uso de las herramientas típicas de la Lógica formal contemporánea.

En el argumento ontológico la existencia de Dios se deduce de la propia idea de Dios y del hecho de que ésta se encuentre en mi mente y sea además libre de contradicción. Si Dios es el ser más perfecto que cabe imaginar y aquello que *existe* muestra mayor perfección que aquello que se ve privado de la existencia, entonces parece necesario que Dios, que goza de todas las perfecciones, exista. Se puede ver claramente de qué modo se pretende deducir a partir del planteamiento del propio problema su solución. ¿Qué modalidad de argumento mirabilis es la que el mentalista aspira a emplear contra la I.A.? Un poco de reflexión puede llevar rápidamente a averiguar algunas de sus características. Se trata de construir un argumento válido y además necesario cuya conclusión sea la existencia de alguna tarea típicamente cognitiva que el ser humano es capaz de ejecutar y no así un ingenio artificial que pretenda imitarlo. Validez y necesidad son dos rasgos característicos de la Lógica, por lo que el argumento debería ser un *argumento lógico*. Debe mostrar la existencia de una tarea al alcance de la mente humana que quede sin embargo fuera de las que puede ejecutar una máquina. Por tanto, lo más sensato perece buscar primero entre aquellas tareas que sabemos que un ingenio mecánico, artificial, no puede ejecutar. El tipo de máquinas que tenemos en mente comparten todas ellas un mismo lenguaje, al punto de hacer prescindible el soporte o el diseño. Pienso en el lenguaje matriz de la Teoría de la Computación que no es otro que aquel en el que se puede construir una máquina de Turing. Algo que es a su vez, sin precisar más ahora, una justa combinación de Lógica y Aritmética elemental. Si somos capaces de hallar alguna limitación relativa a lo que ese lenguaje es capaz de caracterizar como tarea efectiva haciéndolo a través de un argumento lógico válido y necesario parece evidente que habré satisfecho los objetivos mentalistas a entera satisfacción. Muchos han creído ver en los *Teoremas de Limitación de Gödel* o en el *Problema de Parada* para máquinas de Turing instancias plenamente aceptables de este tipo de argumento.

Los Teoremas de Limitación de Gödel -establecidos en 1931- muestran que bajo ciertas condiciones es posible hallar una proposición verdadera de la aritmética que, no obstante, resulta imposible demostrar en ese sistema formal, es decir, en la propia aritmética. Puesto que las máquinas con que ahora trabajamos no son sino sistemas simbólicos parece obvio que el modo en que fabrican outputs es exactamente el mismo por medio del cual la versión formal de la aritmética fabrica sus teoremas. Debemos admitir, por tanto, que ningún ingenio simbólico artificial puede producir aquella proposición que nosotros hemos sido capaces de identificar como verdadera. Este argumento, así visto, parece satisfacer todas las condiciones de contorno que

previamente hemos identificado en el tipo de argumento mirabilis capaz de mostrar la superioridad de la mente sobre cualquier construcción mecánica de tipo simbólico.

El primer uso de este tipo de argumento se encuentra en una obra destinada a explicar el contenido e importancia de los Teoremas de Gödel, de hecho, una de las primeras orientadas a tal fin y seguramente la mejor de todas ellas. Se trata del *Gödel's Proofs* de E. Nagel y J.R. Newman publicado por vez primera en 1958. En su capítulo final estos autores se *deslizan* hacia este tipo de conclusiones sin que parezca haber una previa elaboración o una intención claramente manifiesta. Quizá sea por ello que el exponente tradicional de este argumento sea un texto algo posterior debido esta vez a J.R. Lucas. En 1961 se publica en el número 36 de *Philosophy* un artículo de este autor titulado "Minds, Machines and Gödel" en cuyas primeras líneas se puede leer lo siguiente:

> "Gödel's theorem seems to me to prove that Mechanism is false, that is, that minds cannot be explained as machines. So also has it seemed to many other people: almost every mathematical logician I have put the matter to has confessed to similar thoughts, but has felt reluctant to commit himself definitely until he could see the whole argument set out, with all objections fully stated and properly met. This I attempt to do."

A esta obra seguirían un cierto número de réplicas alentadas en buena medida por la considerable polémica provocada por el artículo original. El debate abierto entonces pareció quedar zanjado después de un severo escrutinio en el que llegó a intervenir el propio Gödel y que puso de manifiesto la incorrecta lectura de sus teoremas hecha por Lucas en su propuesta antimecanicista. No entraré ahora en el detalle de la *corrección* del argumento mirablis de Lucas ya que ocuparía el resto de este escrito y provocaría, eso me temo, más dudas que certezas. Lo cierto es que pese al amplio consenso obtenido entonces por la postura contraria a admitir los argumentos de Lucas este debate no ha quedado del todo cerrado. Los detractores del mecanicismo –como en ocasiones se denomina al proyecto de la I.A. en su acepción más amplia- nunca quedaron del todo convencidos de las respuestas procedentes del entorno de la Lógica formal quizá en exceso técnicas y poco eficaces en ocasiones desde el punto de vista intuitivo. Este hecho pudo contribuir de forma decisiva a la preservación de un cierto estado de ánimo capaz de retornar con nuevo impulso en la obra que publica Penrose en 1989 y que lleva por título *The Emperor's New Mind: Concerning Computers, Minds, and The Laws of Physics*. Este trabajo reproduce de forma prácticamente idéntica los propios argumentos de Lucas sin que parezca existir una relación directa ni tampoco un ánimo de plagio. Se trata, o eso me parece, de un genuino reencuentro con una tesis que de un modo u otro había quedado larvada en un debate aún abierto. La cascada de réplicas y contrarréplicas involucra de nuevo a lo más granado de la comunidad científica –véase el monográfico de la revista Psyche, (Psyche vol.2)- y vuelve a señalar las mismas debilidades que ya en su día fueran indicadas a propósito de la obra de Lucas. La respuesta de Penrose, seguramente más capaz desde un punto de vista técnico que el propio Lucas, se sustancia en otra obra publicada en 1994 y titulada esta vez *Shadows of the Mind: A Search for the Missing Science of Consciousness*. La novedad que se aprecia en esta ocasión tiene que ver con el punto de partida, es decir, con el argumento que se toma como modelo de tarea o acto cognitivo que un ingenio mecánico no puede reproducir en modo alguno. En

esta ocasión Penrose se cuela en la propia alcoba de la Teoría de la Computación para obtener el secreto que le ha de permitir fijar sus objetivos antimecanicistas. Porque, ¿qué mejor punto de partida que aquel que establece una limitación referida a las propias máquinas de Turing, es decir, referida al marco teórico en el que se sustenta la propia I.A.?

El resultado al que me refiero es hoy conocido como *Problema de Parada –halting problem-* y es planteado por primera vez en el artículo seminal de Turing de 1936-Turing no emplea sin embargo esos términos, que son introducidos años más tarde por Martin Davis en el clásico de Teoría de la Computación que publica en 1958-. El Problema de Parada muestra la imposibilidad de resolver de forma mecánica el problema consistente en determinar si una cierta máquina de Turing que actúa sobre un input determinado alcanzará un estado de parada arrojando el output correspondiente. Dicho de otra forma, no podemos diseñar una máquina de Turing que responda de forma general a este problema para cualquier máquina e input dados. El argumento de Penrose parece ahora bastante evidente. Es obvio que podemos diseñar máquinas que arrojen como output un cierto valor previamente fijado cuando una máquina dada finaliza su rutina arrojando un output, en ello no hay problema. El problema, sostiene Penrose, surge cuando comprobamos que mediante un razonamiento lógico externo a la rutina de la propia máquina podemos establecer que ésta no para arrojando un output mientras que ella misma sólo es capaz de continuar su rutina sin ser capaz de alcanzar la misma conclusión a la que nosotros hemos llegado. Así las cosas, el argumento parece bastante robusto. Pero al igual que en el resto de los casos podemos llegar a la conclusión, tras un severo y delicado escrutinio, eso sí, de que reposa sobre consideraciones dudosas y pasos inferenciales no justificados. Su estructura lógica es, por decirlo de algún modo, deficiente.

Como se puede ver, no estamos hablando de viejos argumentos o de polémicas interesantes sólo para el historiador, sino de algo extraordinariamente próximo en el tiempo y en el contenido. Pero, ¿qué futuro cabe atribuir a este tipo de maniobras? No tengo especial problema en confesar mi profundo escepticismo ante el posible éxito de algún tipo de argumento mirabilis, pero tampoco puedo vaticinar aquí y ahora su definitiva derrota. Es más, puedo atreverme, y creo que no corro mucho riesgo en ello, a pronosticar nuevas réplicas de mayor o menor cuantía. Que manifieste mi escepticismo ante este género argumentativo no debe interpretarse como un brindis a favor de las posibilidades del paradigma mecanicista y por tanto de la I.A. No hay ninguna necesidad en ello. Gödel, antimecanicista declarado, nunca aceptó este tipo de maniobras basadas en sus propios teoremas aunque sí las practicó en lo que se refiere al problema de la existencia de Dios –como ya he dicho Gödel llegó a elaborar una variante del argumento ontológico relativamente poco conocida-. Aunque resulte paradójico, creo que hemos aprendido mucho más de nuestra profesión e intereses intentado desmontar este tipo de argumentos que lo que hubiéramos conseguido en su ausencia. Lo único que pretendo dejar claro aquí es la conveniencia de prestarles la debida atención evitando en la medida de lo posible actitudes dogmáticas basadas en el prejuicio. Como se puede ver, no es cosa de *filósofos* o, en general, de investigadores ajenos al *arte*. Se trata de intentos genuinos de extraer de las condiciones de un problema los términos de su solución y de paso la propia solución. Pensar que siempre hay en esto algo de exceso es una postura con la que ciertamente simpatizo. Pero tampoco me atrevo a rechazar cualquier intento en esa dirección.

Porque como sucede en el ámbito de la guerra, en el que la diferencia entre la insensatez y la valentía suele ser el éxito, la diferencia entre una genuina demostración y un argumento mirabilis puede ser también cuestión de éxito y no de método.

4. Las alternativas al modelo clásico.

¿Por qué creo que un cierto conocimiento del trasfondo formal de la Teoría de la Computación puede ayudar a entender mejor las posibles alternativas al modelo computacional vigente? Simplemente porque permite entender de manera razonada en qué consiste una de esas posibles alternativas y con qué se comprometen aquellos que dicen estar en posesión de alguna de ellas.

Presentar un modelo computacional alternativo al modelo clásico no consiste tan solo en elaborar un marco conceptual muy alejado del que hoy representan las máquinas de Turing, por ejemplo. Las redes neuronales fueron consideradas de forma ingenua como una alternativa en este sentido hasta que se demostró que toda tarea que pudiera ser implementada a través de una de estas redes también podía ser programada en términos de una máquina de Turing y por tanto reducida a computación estándar. No quiero decir que las diferencias de planteamiento no sean en ningún caso relevantes y mucho menos en casos tan extremos como los que acabo de mencionar. Lo que en realidad sostengo es la necesidad de tomarse en serio el carácter formal y simbólico de la computación teórica. Expresar computacionalmente una cierta tarea no es sino ofrecer su definición en un lenguaje simbólico apropiado. Existen infinitas formas de realizar esta exigencia, pero nada garantiza que una de esas maneras de traducir el carácter computacional o efectivo de una rutina sea realmente capaz de programar alguna tarea que resulte inaccesible al resto de los lenguajes de programación existentes o imaginables.

La comparación de los recursos expresivos de dos lenguajes de programación distintos es una tarea típica dentro del dominio de la lógica formal o lo que es lo mismo en este caso, de la computación teórica. No siempre es fácil establecer uno de estos resultados, pero el *modus operandi* sí resulta sencillo de entender. Para mostrar que un lenguaje determinado es capaz de programar exactamente las mismas rutinas que otro dado, basta tomar cada uno de los recursos disponibles en el primero de ellos y establecer su correspondencia exacta en el segundo. Este procedimiento, si finaliza con éxito, permite traducir cualquier programa expresado en el lenguaje original a otro del lenguaje sobre el cual se realiza la comparación. Para establecer la plena equivalencia entre ambos lenguajes basta repetir el proceso invirtiendo ahora el lenguaje tomado como punto de partida.

Este tipo de teoremas tuvo su importancia en los primeros momentos de la Teoría de la Computación debido a la simultánea coexistencia de una serie de modelos cuya equivalencia no estaba probada y la cual distaba además de ser evidente. Este hecho, a menudo olvidado u omitido, es determinante para el curso que después han tomado los acontecimientos. Porque, ¿qué hubiera impedido que existieran modelos alternativos no comparables entre sí? Realmente no lo sabemos. Lo cierto es que ese y

no otro era el resultado más esperable en una época –mediados de la década de 1930- bastante acostumbrada a las muestras de mala conducta por parte de las ciencias formales, Lógica y Matemáticas. Pero resultó que los tres principales modelos propuestos de forma casi simultánea –no entraré ahora en detalles históricos- la *teoría de las funciones recursivas generales* de Gödel, el *cálculo lambda* de Church y las *máquinas de Turing* resultaron equivalentes en el sentido que acabo de detallar. Años más tarde Gödel llegaría a confesar su sorpresa ante la sorprendente estabilidad mostrada por un concepto, el de *calculabilidad efectiva* que, en el fondo, sólo procede de nuestra intuición más general. Y para el cual no parece haber alternativas formales en el sentido que me he preocupado en aclarar.

Church se ocuparía de formular esa impresión general en términos de una tesis –hipótesis en realidad- que hoy conocemos como *Tesis de Church* y según la cual *toda tarea efectiva puede ser expresada en el formalismo de las máquinas de Turing, el cálculo lambda, etc*. Es importante entender que no se trata de un *teorema* establecido por alguno de esos sutiles argumentos que los lógicos denominamos *pruebas*, sino de una tesis tan solo apoyada –que no es poco- en 70 años de experiencia. Durante todo este tiempo nadie ha llegado a establecer un sistema simbólico capaz de ejecutar una tarea de forma efectiva que no pudiera ser programada igualmente en términos de máquinas de Turing, es decir, por procedimientos estándar.

Se puede entender por qué no resulta fácil, aunque sí tentador, arriesgarse a proponer algún modelo alternativo: el riesgo de fracaso es, simplemente, máximo.

De todo lo que se ha podido decir a propósito de los intentos por diseñar modelos alternativos me sigue pareciendo que lo más sutil es lo que en su día pudo dar a entender Gödel cuando fue preguntado al respecto por su biógrafo intelectual, el también lógico Hao Wang. Su sugerencia, sólo apuntada en sus palabras, dirige nuestra mirada a las condiciones formales en que es definido el simbolismo de las máquinas de Turing. Si realmente hemos de obtener un modelo alternativo éste tiene que diferir de entrada en alguno de esos componentes. Es decir, tiene que haber algo que actúe ya desde un principio de forma realmente distinta.

De los intentos propuestos recientemente los hay que parecen haber prestado atención a este consejo y también otros que se han ido alejando del mismo. La computación cuántica parecía prometer un marco realmente distinto aunque sólo fuera debido a que la lógica que subyace a los fenómenos de tipo cuántico no es la estándar. El tiempo ha ido rebajando esas expectativas en la medida en que los lenguajes de programación cuánticos no han sido capaces de establecer, hasta la fecha, resultados en la línea de lo esperado.

El caso que me parece más merecedor de atención nos sitúa en un plano teórico ciertamente abstracto. Me refiero al tipo de propuestas que se reúnen bajo el enunciado genérico de *supertareas*. Existe una larga tradición filosófica alrededor de este concepto cuyo ancestro más remoto es la propia paradoja de Aquiles y la Tortuga. Pero, ¿es posible programar supertareas? Imaginemos una máquina de Turing cualquiera a la que hemos acoplado una fuente de energía capaz de hacer que cada paso se de en la mitad de tiempo que el anterior. Supongamos que en ejecutar su primera instrucción nuestra máquina invierte 0.5'', por tanto, el segundo paso no le supondrá más de 0.25''. No hace falta saber muchas matemáticas para convencerse de que esta peculiar máquina siempre habrá terminado su rutina, sea la que sea, y

suponga el número de pasos que suponga –incluyendo el caso el caso en que estos son infinitos- en 1'' –la serie $\sum 1/n^2$ es convergente-. Estas máquinas reciben en la actualidad el nombre de *máquinas acelerantes* –Acelerating Turing Machines- y vienen asociadas, aunque no en exclusiva al nombre de J. Copeland-. No creo que sea necesario aclarar que en la actualidad nadie ha conseguido implementar una de estas máquinas. Lo que quizá sí sea bueno advertir es sobre la existencia de un encendido debate acerca de si el propio concepto está o no exento de problemas e incluso contradicciones. Dejaré este problema para aquellos que deseen entrar en ello.

Las máquinas acelerantes presentan una ventaja sustancial frente a sus colegas estándar por la sencilla razón de que resuelven –y de forma trivial- el problema de parada de estas últimas. Obsérvese que cuando la máquina estándar de Turing computa un input para el cual no existe un output lo único que ésta hace es continuar sus cálculos sin término aparente. La variante acelerante termina a tiempo –en el segundo canónico- dejando la cinta de cálculo en blanco. Esta clase de recursos, si bien es simbólicamente idéntica a la estándar, parece capaz de computar tareas que no están al alcance de la clase de las propias máquinas de Turing. Curiosa conclusión para todos aquellos que piensan que la computación es *sólo* un asunto del lenguaje simbólico en el que se opera.

Sea como fuere hay aún muchas conclusiones que obtener de este tipo de propuestas, entre otras, las propias condiciones de clausura de la clase de las funciones Turing-acelerantes y en general todos los rasgos básicos de este tipo de entidades. Y ello antes, sin duda, de que podamos pensar en la identificación de algún tipo de proceso natural capaz de implementar una de estas supertareas.

No quiero que parezca que esta hipótesis me es más simpática que otras. Si la he desarrollado con algo más de detalle es para que se entienda el tipo de análisis crítico del concepto clásico que he pretendido deducir de las palabras de Gödel. Este es un caso en el que ciertamente nos alejamos y mucho del modo de operar de las máquinas de Turing. La cuestión es saber si estamos dispuestos a pagar el precio que de ello se deriva.

Hay muchas otras iniciativas que de un modo u otro van en la misma dirección. Las *Infinite Time Turing Machines* –ITTM- de Hamkin y Lewis pueden considerarse, por ejemplo, como un desarrollo más general y abstracto también en la línea de las máquinas acelerantes y algo menos interesante, en mi opinión, en la medida en que eliminan el comportamiento temporal de éstas últimas. El *pi-calculus* desarrollado por Milner puede ser entendido como un intento, aunque bastante sofisticado, de desarrollar la idea de interacción dentro de un sistema computacional complejo. Intuición ciertamente desatendida en el modelo clásico. Los sistemas paralelos asíncronos podrían ser incluidos también en este apartado y la lista seguiría añadiendo variantes más o menos pertinentes por tiempo aún indefinido.

No sé cuánto cabe esperar de este tipo de iniciativas que, en el mejor de los casos, son sólo propuestas teóricas no exentas de problemas. Pero, ¿acaso no es esa la forma de actuar y progresar de esta disciplina?

5. A vueltas con los experimentos mentales.

La Inteligencia artificial no es un proyecto que se siga necesariamente de los planteamientos adoptados por la Teoría de la Computación durante los años 30 y 40. Hacen falta más cosas para que podamos plantearnos realmente la posibilidad de fabricar inteligencia no humana. Los ingredientes que es preciso añadir para obtener esta mezcla proceden de la formulación de un experimento mental que con el tiempo ha llegado a obtener una cierta fama. Me refiero, como no puede ser de otra manera al conocido como Test de Turing –cuyo nombre original es *juego de la imitación*-. Este experimento mental fue elaborado por Turing en 1950 en un artículo titulado *Computing Machinery and Intelligence* que aparece en el número 49 de la revista *Mind*. Sus primeras palabras son suficientemente expresivas:

> "I PROPOSE to consider the question, 'Can machines think?'"

El detalle podemos dejarlo a un lado –por conocido- y aclarar cuál es la sustancia del problema. ¿Qué nos hace merecedores de los rasgos de la humanidad y con ello la inteligencia? Pocas cuestiones pueden tener un mayor gusto filosófico que la que con toda intención acabo de plantear. Porque no es ésta una pregunta que tenga mucho sentido hacer. Todos sabemos en qué consiste nuestra humanidad. Bien, pero si es así, ¿por qué tanta dificultad en encontrar una respuesta? Esta es la grieta por la que penetra el argumento construido por Turing con una finura y eficacia realmente sorprendentes. Para decir de alguien que posee los rasgos de la humanidad sólo hay, afirma Turing, un recurso admisible: su conducta. Si una entidad demuestra, bajo cualesquiera condiciones imaginables, una *conducta* en todo idéntica a aquella que es propia de los seres humanos, ¿qué razones podríamos tener para negarle la condición humana? ¿El aspecto, el material del que está hecho? Admitir que dos entidades *funcionalmente* equivalentes son la misma entidad, o que pertenecen en general a la misma clase, constituye el credo de una escuela de pensamiento conocida, por ello mismo, como *funcionalismo*. Sutilezas aparte, el funcionalismo, en esta u otra acepción técnicamente más ajustada, constituye la raíz del argumento de Turing y con ello del programa amplio de la I.A.

Esta versión del funcionalismo participa a su vez de un fuerte regusto empírico en la medida en que reconoce que el modo de comparar entidades funcionales en aquellos casos en que no se puede acceder a algo así como su diseño, es mediante la conducta manifiesta.

La gran virtud del experimento de Turing fue el ser capaz de invertir lo que los juristas denominan la *carga de la prueba*, situando a los partidarios del mentalismo en una posición ciertamente incómoda. Hasta ese momento abrazar la posibilidad de fabricar inteligencia no humana llevaba a sus defensores a comprometerse con alguna demostración práctica simplemente como medio de admitir la posibilidad general del proyecto. Me explicaré con algo más de cuidado. Antes de la propuesta de Turing no existían razones teóricas para admitir siquiera la posibilidad de producir cognición artificial, salvo en un sentido muy general parecido al que otorgamos a afirmaciones del tipo "es posible que mañana sea el fin del mundo". La única forma de otorgar plausibilidad a esa afirmación era mediante una maniobra ciertamente extrema:

fabricando de hecho inteligencia artificial. El Test de Turing invierte la carga de la prueba al mostrar que aquellos que niegan en principio esta posibilidad asumen, al proceder así, creencias poco racionales acerca de aquello en que consiste la conducta humana. Turing hizo intelectualmente respetable el proyecto de la I.A., ese es su gran mérito.

Pero si algo enseña la historia es que un experimento mental a menudo sólo es compensado con otro de similar efecto. Esto es lo que sucedió en 1980 cuando J. Searle formula en "Minds, Brains, and Programs.", artículo publicado en el número 3 de *Behavioral and Brain Sciences*, el experimento de la *habitación china*. Es posible que esta fecha marque una especie de punto de inflexión en el progreso de la I.A. pasando de una etapa feliz a otra más madura en la que el progreso empieza a ralentizarse y las dificultades se muestran ya como algo evidente.

Este experimento es propuesto 30 años después de que Turing formulara el suyo, dato que creo que conviene tener muy en cuenta. Y se construye como una forma de responder a la pretendida ofensiva del modelo de lo que Searle denomina la I.A. fuerte. Es decir, aquella basada en el paradigma computacional y por tanto de filiación funcionalista. Añado este dato porque Searle no niega la posibilidad de inteligencia artificial, pero rechaza la posibilidad de que ésta pueda obtenerse a través de entidades simbólicas sin otro tipo de propiedades causales como las que el cerebro humano posee.

La habitación china no es sino una metáfora útil para evaluar el modo en que la ejecución de una rutina determinada implica o supone cognición. La opinión de Searle al respecto es claramente negativa. Si en una sala ubicamos a un sujeto que *transcribe*, siguiendo instrucciones antes escritas en un papel, ideogramas chinos al castellano, ¿debemos afirmar que este sujeto *sabe* chino? La conclusión resulta obviamente negativa. Del mismo modo que este sujeto no sabe chino por el mero hecho de aplicar unas instrucciones determinadas, un programa no adquiere o desarrolla cognición por mucho que su conducta así lo pueda dar a entender.

Se ha hablado mucho sobre este experimento y sobre la forma en que afecta al proyecto de la I.A. y no está mi ánimo entrar ahora en un debate que ocupa estanterías enteras en las bibliotecas especializadas. Sea como fuere lo que sí puede apuntarse en su haber es un logro muy similar al que en su día consiguiera para sí el Test de Turing. Ha conseguido mostrar los excesos y los compromisos que una actitud funcionalista ingenua puede llegar a provocar. Defender la I.A. después de Searle, no es, admitámoslo, una tarea tan fácil e inocente como lo era antes de su decisiva contribución.

Es posible que este tipo de experimentos no arrojen por sí solos resultado alguno de valor para lo que nos ocupa. Pero tampoco habría que juzgarlos por algo que en ningún momento se proponen. Se trata en todos los casos de mostrar, a través de ficciones teóricamente plausibles y pertinentes, las consecuencias o posibilidades de una cierta hipótesis. El impacto de este tipo de argumentos es muy variable y depende en buena medida de la sensibilidad de las personas y del grado de receptividad a ideas procedentes de otros campos. Lo que es indudable en cualquier caso es la capacidad de estos experimentos para producir, lentamente si se quiere, estados de ánimo determinantes en muchos casos para el progreso o estancamiento de una tesis o posición. Este efecto se logra, como hemos visto, a través de la presentación de una serie de hitos necesitados de respuesta o tratamiento y sin los cuales no parece posible

avanzar en la dirección deseada. Tienen la capacidad de volverse tópicos dentro del tratamiento de un problema haciéndonos circular por ciertos caminos y no por otros. Ignorarlos puede ser fatal para el estado de la cuestión en una disciplina. Estimular su debate e incluso animar a la propuesta de experimentos de este tipo en épocas de relativo impasse –como la presente- puede ser no sólo bueno para estimular nuevas ideas, sino a veces lo único que se puede hacer.

6 De visita en el Museo de Historia Natural de las Ideas.

Cuando comunidades científicas fuertemente profesionalizadas admiten y reclaman en sus salones la presencia del filósofo quizá sea bueno empezar a interesarse por el estado de salud de proyectos. Porque, admitámoslo, no se suele recurrir a colaboraciones tan comprometidas cuando todo funciona como es debido. Invitar al otro, al extraño es, sin embargo, uno de los mejores remedios que conozco al vicio del ensimismamiento. Y el ensimismamiento suele ser causa de decadencia.

Como ya dije al comienzo de este texto, la visita al Museo de Historia Natural de las Ideas de la que puedo hablar es más bien la mía y no la que cabría esperar de un guía autorizado. Se trata de un recorrido limitado a unas pocas salas a las que he ido volviendo según mis intereses evolucionaban. Son solo ideas a las que con el tiempo he ido dedicando cada vez mayor atención y he aprendido a valorar como algo potencialmente relacionado con mis intereses en el ámbito de la Computación Teórica y de la propia Filosofía. Sólo las mencionaré, apenas nada más.

La forma clásica de entender el fenómeno de la conciencia hace de ésta una entidad que sobreviene a partir de una cierta etapa en el desarrollo evolutivo del cerebro. Y lo hace en términos de todo o nada. La conciencia se interpreta pues como un resultado de la inteligencia. Sin inteligencia no hay conciencia, aunque nada garantiza que la primera baste para hacer emerger la segunda. La etología, y sobre todo su aplicación al tratamiento de la conducta de los grandes primates ayuda a imaginar otras posibilidades tentadoras por su capacidad para sacarnos de esquemas quizá demasiado gastados. ¿Podríamos imaginar la conciencia como un hecho independiente de la inteligencia? Es decir, como un fenómeno cognitivo autónomo con sus propias reglas. ¿Podríamos imaginar la conciencia no como el resultado de un cierto modo de procesar información, sino como una *peculiar* forma de procesar información? ¿Podríamos averiguar cuáles son los hechos simples de la conciencia, determinar su código, el lenguaje en que opera, del mismo modo que en su día fuimos capaces de obtener el código de los cálculos que hacemos de manera efectiva? Como digo son sólo elementos para animarnos a transitar por caminos aún poco explorados pero que corren en paralelo a líneas de investigación animadas desde la etología y que se resisten a negar a otros seres vivos no tan distintos a nosotros algún tipo de conciencia como aquella que orgullosamente nos atribuimos de forma exclusiva.

Siguiendo esta dirección en la que me cuestiono, como principio metodológico, eso sí, la supremacía del ser humano, podemos encontrar otra pieza notable esta vez en el dominio de la paleontología. Nuestra especie no ha sido la única que ha habitado esta

Tierra y a la cual cabe atribuir inteligencia consciente. Parece probado que ha habido al menos otras tres formas distintas de entender la humanidad. *Erectus* era humano y mostraba, no sólo habilidad técnica, sino también respeto por los miembros del clan y lo que podría ser un incipiente sentido de la trascendencia y el yo. Los neandethales y nosotros mismos hemos compartido territorio y quién sabe si también otras pasiones durante un prologando periodo de tiempo. Todos somos humanos, pero probablemente de modos muy distintos. Es una lástima que nunca podamos llegar a saber con certeza en qué consistieron esas diferencias. ¿Por qué hablamos entonces de *inteligencia artificial* como un asunto de todo o nada entendido sin matices y en única correspondencia con el modo de actuar que caracteriza a la única especie de homo que hoy queda en la Tierra? Admitir que la inteligencia es un fenómeno que puede darse de modos distintos resulta más fácil de digerir como hipótesis una vez que hemos reconocido que la propia humanidad es un hecho que admite más de una lectura posible. Una vez alcanzado ese punto es fácil dar el salto que le acompaña. Si la inteligencia puede darse de formas que quizá aún no hemos descrito, ¿qué queda del proyecto de la I.A.? Precisamente el hecho de la *artificialidad*. La posibilidad de generar entidades capaces de alcanzar conductas independientes y autónomas como aquellas que caracterizan a los seres vivos. Pienso, para no prolongar el suspense, en el fenómeno que en biología se conoce como *especiación*. La especiación describe el momento en que los procesos de selección natural consiguen aislar un sistema genético limitando el intercambio de genes a entidades del mismo grupo. Tengo la impresión, incluso estaría dispuesto a apostar por ello, de que en el futuro habremos de considerar muy en serio algún comportamiento capaz de sorprendernos por su relativa autonomía. La autonomía de una entidad del tipo que sea sólo nos debe preocupar cuando afecte a los mecanismos de especiación capaces de perpetuarla de forma sistemática. La especiación será entonces el problema, y no la inteligencia.

Estas consideraciones llevan a las siguientes sin apenas solución de continuidad. El marco teórico sobre el que ha trabajado la I.A. en décadas pasadas –esto ya no es cierto en la actualidad- depende en gran medida del modelo de procesamiento simbólico encarnado en la estructura de un antiguo ordenador personal –PC-. Es posible que esto no sea apreciable para el gran público, pero nuestros actuales ordenadores de sobremesa, si bien aún son *personales*¸ ya no son PC's. Son entidades esencialmente asociadas a la Red. No pueden sobrevivir por mucho tiempo desconectados. ¿A qué voy con todo esto? Simplemente a reivindicar el carácter eminentemente social de la inteligencia. Concebir modelos de nuestra actividad cognitiva aislados en el entorno artificial que representa un pequeño –o gran- ordenador de sobremensa supone una interpretación completamente equivocada de la inteligencia humana. Nuestra capacidad simbólica, lingüística y lógica es ante todo una solución adaptiva entre otras posibles y una muy costosa por cierto. Para obtener inteligencia artificial, ¿no deberíamos considerar en serio el tipo de entornos y presiones que fomentaron nuestro progreso evolutivo? Hace tiempo que los algoritmos genéticos desarrollan ideas similares aunque en contextos más limitados. La aparición de los sistemas multiagente parecen moverse en una dirección similar al tomar en serio la idea de sistema complejo a la hora de definir conducta. Desde el ámbito de la filosofía podríamos hablar del interés por el concepto de *cognición extendida* como iniciativa afín. Según los defensores de esta propuesta los actos cognitivos del ser humano no se pueden explicar adecuadamente en ausencia de un

medio sobre el que nuestra mente pueda operar manipulándolo físicamente a través de nuestros miembros. Todas estas corrientes parecen compartir una cierta incomodidad ante las interpretaciones solipsistas de la mente y la inteligencia. No puedo negar mi simpatía ante esta forma de reinterpretar nuestros objetivos. Las preocupaciones que quizá pasen a ocuparnos en los próximos años tendrán por tanto más que ver con el medio, las condiciones y la medida de las presiones selectivas en que debe operar una entidad supuestamente inteligente que con la propia definición de concepto tan resbaladizo. Pero todo tiene sus riesgos, porque si nos tomamos en serio el hecho de que la inteligencia es el producto de la evolución, ¿qué garantiza que por medios artificiales vayamos a ser capaces de crear las tensiones necesarias en menos tiempo que el que la propia naturaleza tuvo que invertir en ello? ¿Y que garantiza que el resultado nos guste o quede convenientemente sometido a nuestro control?

Confieso que aquí más que el filósofo lo que asoma es la vieja pasión del aficionado a la literatura fantástica, pero lo cierto es que siempre he encontrado inspiración y consuelo en aquellas narraciones que imaginan sin vergüenza aquello que nosotros sólo nos atrevemos a pensar y confesar en la intimidad de nuestro gabinete. Mucho me temo que nuestras mejores opciones hoy en día no pasan por contribuir a las líneas que otros desarrollan hace años con mayor o menor éxito, sino en atrevernos a poner un pie donde ellos aún no han llegado.

Mi visita termina aquí. He intentado mostrar de forma panorámica, a través de meros ejemplos en ocasiones, lo que la Filosofía y la Lógica pueden hacer en ese diálogo entre personas e intereses que se supone ha de producirse en un marco interdisciplinar. Ahora me toca a mi escuchar a los demás.

Towards Artificial Creativity: Examples of some applications of AI to music performance

Ramón LOPEZ DE MANTARAS
IIIA, Artificial Intelligence Research Institute CSIC, Spanish Scientific Research Council CampusUAB08193 Bellaterra, Spain

Abstract. Creativity is an important element of our problem solving capabilities that involves, among others, heuristic search, reasoning, analogy, learning, and reasoning under constraints. In this paper I describe examples of computer programs capable of replicating some aspects of creative behavior in the domain of music

Keywords. Artificial Intelligence, Creativity, Music

1 Introduction

New technologies, and in particular Artificial Intelligence, are drastically changing the creative processes. Computers are playing very significant roles in creative activities such as music, architecture, fine arts and science. Indeed, the computer is already a canvas, a brush, a musical instrument, etc. However, I believe that we must aim at more ambitious relations between computers and creativity. I mean relations other than just seeing the computer as a tool to help human creators but a creative entity in itself.

In this paper I will, therefore, briefly address the question of the possibility of building creative machines. But, what is creativity? Why is it so mysterious?. Creativity seems mysterious because when we have creative ideas it is very difficult to explain how we got them and the best we can do is talk about "inspiration", "intuition" when we try to explain creativity. The fact that we are not conscious about how a creative idea manifests itself does not necessarily imply that a scientific explanation cannot exist. As a matter of fact, we are neither aware of how we perform other activities such as language understanding, pattern recognition, etc. but we have better and better AI techniques able to replicate such activities.

For many, creativity requires the possession of a mysterious and unusual "gift" that cannot be explained. This is possibly due to the fact that, often, creativity is associated with geniuses like, Bach or Picasso, but should the term "creative" be only applied to a few men and women? Or is it rather one more aspect of intelligence? In other words, is creativity the result of a very special mechanism only existing in very unusual minds or is rather one more aspect of our problem solving activity? An

A. Fernández-Caballero & S. Miguel (Eds.): 50 Años de la Inteligencia Artificial, pp. 43-49, 2006.

operational, and widely accepted, definition of creativity is "a creative idea is a novel and valuable combination of known ideas". In other words, new physical laws, new theorems, new musical pieces can be generated from a finite set of existing elements. I agree with this definition and, therefore, with those that believe that creativity is an advanced form of our problem solving capabilities that involves, among others, memory, heuristic search, reasoning by analogy, learning and reasoning under constraints and therefore it should be possible to replicate by means of computers.

Based on this belief, in this paper I give some examples of representative applications to expressive music performance that employ AI techniques.

For further reading I recommend the books by Boden [1,2], Dartnall [3], Partridge & Rowe [4], and Bentley & Corne [5]; as well as the papers by Rowe & Partridge [6], and Buchanan [7]. Hay veces en las que el reparto de tareas representa una buena medida a la hora de abordar un problema. De hecho, las más. Hay situaciones en las que incluso no se puede proceder de otro modo, pero plantearse este logro como punto de partida se parece demasiado al reparto de distritos y regiones que podemos recordar de clásicos del cine dedicados al mundo del *hampa*.

2 Synthesizing expressive music

One of the main limitations of computer-generated music has been its lack of expressiveness, that is, lack of "gesture". Gesture is what musicians call the nuances of performance that are unique and subtly interpretive or, in other words, creative.

One of the first attempts to address expressiveness in music is that of Johnson [8]. She developed an expert system to determine the tempo and the articulation to be applied when playing Bach's fugues from "The Well-Tempered Clavier". The rules were obtained from two expert human performers. The output gives the base tempo value and a list of performance instructions on notes duration and articulation that should be followed by a human player. The results very much coincide with the instructions given in well known commented editions of "The Well-Tempered Clavier". The main limitation of this system is its lack of generality because it only works well for fugues written on a 4/4 meter. For different meters, the rules should be different. Another obvious consequence of this lack of generality is that the rules are only applicable to Bach fugues.

The work of the KTH group from Stockholm [9, 10, 11, 12] is one of the best known long term efforts on performance systems. Their current "Director Musices" system incorporates rules for tempo, dynamic, and articulation transformations constrained to MIDI. These rules are inferred both from theoretical musical knowledge and experimentally by training, specially using the so-called analysis-by-synthesis approach. The rules are divided in three main classes: Differentiation rules, which enhance the differences between scale tones; Grouping rules, which show what

tones belong together; and Ensemble rules, that synchronize the various voices in an ensemble.

Canazza et al [13] developed a system to analyze how the musician's expressive intentions are reflected in the performance. The analysis reveals two different expressive dimensions: one related to the energy (dynamics) and the other one related to the kinetics (rubato) of the piece. The authors also developed a program for generating expressive performances according to these two dimensions.

The work of Dannenberg and Derenyi [14] is also a good example of articulation transformations using manually constructed rules. They developed a trumpet synthesizer that combines a physical model with a performance model. The goal of the performance model is to generate control information for the physical model by means of a collection of rules manually extracted from the analysis of a collection of controlled recordings of human performance.

Another approach taken for performing tempo and dynamics transformation is the use of neural network techniques. In [15] a system that combines symbolic decision rules with neural networks is implemented for simulating the style of real piano performers. The outputs of the neural networks express time and loudness deviations. These neural networks extend the standard feed-forward network trained with the back propagation algorithm with feedback connections from the output neurons to the input neurons.

We can see that, except for the work of the KTH group that considers three expressive resources, the other systems are limited to two resources such as rubato and dynamics, or rubato and articulation. This limitation has to do with the use of rules. Indeed, the main problem with the rule-based approaches is that it is very difficult to find rules general enough to capture the variety present in different performances of the same piece by the same musician and even the variety within a single performance [16]. Furthermore, the different expressive resources interact with each other. That is, the rules for dynamics alone change when rubato is also taken into account. Obviously, due to this interdependency, the more expressive resources one tries to model, the more difficult is finding the appropriate rules.

We have developed a system called SAXEX that won the Best Paper Award at the 1997 International Computer Music Conference [17]. SAXEX is a computer program capable of synthesizing high quality expressive tenor sax solo performances of jazz ballads based on cases representing human solo performances. Previous rule-based approaches to that problem could not deal with more than two expressive parameters (such as dynamics and rubato) because it is too difficult to find rules general enough to capture the variety present in expressive performances. Besides, the different expressive parameters interact with each other making it even more difficult to find appropriate rules taking into account these interactions.

With CBR, we have shown that it is possible to deal with the five most important expressive parameters: dynamics, rubato, vibrato, articulation, and attack of the notes. To do so, SaxEx uses a case memory containing examples of human performances, analyzed by means of spectral modeling techniques and background musical knowledge. The score of the piece to be performed is also provided to the system. The heart of the method is to analyze each input note determining (by means of the background musical knowledge) its role in the musical phrase it belongs to, identify and retrieve (from the case-base of human performances) notes with similar roles, and

finally, transform the input note so that its expressive properties (dynamics, rubato, vibrato, articulation, and attack) match those of the most similar retrieved note. Each note in the case base is annotated with its role in the musical phrase it belong to as well as with its expressive values. Furthermore, cases do not contain just information on each single note but they include contextual knowledge at the phrase level. Therefore, cases in this system have a complex object-centered representation.

Although limited to monophonic performances, the results are very convincing and demonstrate that CBR is a very powerful methodology to directly use the knowledge of a human performer that is implicit in her playing examples rather than trying to make this knowledge explicit by means of rules. Some audio results can be listen at www.iiia.csic.es/arcos/noos/Demos/Aff-Example.html. More recent papers by Arcos and Lopez de Mantaras [18] and by Lopez de Mantaras and Arcos [19], describe this system in great detail.

Based on the work on SaxEx, we have decelopped TempoExpress [20] a casebased reasoning system for applying musically acceptable tempo transformations to monophonic audio recordings of musical performances. TempoExpress has a rich description of the musical expressivity of the performances, that includes not only timing deviations of performed score notes, but also represents more rigorous kinds of expressivity such as note ornamentation, consolidation, and fragmentation. Within the tempo transformation process, the expressivity of the performance is adjusted in such a way that the result sounds natural for the new tempo. A case base of previously performed melodies is used to infer the appropriate expressivity. The problem of changing the tempo of a musical performance is not as trivial as it may seem because it involves a lot of musical knowledge and creative thinking. Indeed, when a musician performs a musical piece at different tempos the performances are not just time-scaled versions of each other (as if the same performance were played back at diferent speeds). Together with the changes of tempo, variations in musical expression are made [21]. Such variations do not only affect the timing of the notes, but can also involve for example the addition or deletion of ornamentations, or the consolidation/fragmentation of notes. Apart from the tempo, other domain specific factors seem to play an important role in the way a melody is performed, such as meter, and phrase structure. Tempo transformation is one of the audio post-processing tasks manually done in audio-labs. Automatizing this process may, therefore, be of industrial interest.

Other applications of CBR to expressive music are those of Suzuki et al. [22], and those of Tobudic and Widmer [23, 24]. Suzuki et al. [22], also use examples cases of expressive performances to generate multiple performances of a given piece with varying musical expression, however they deal only with two expressive parameters. Tobudic and Widmer [23] apply instance-based learning (IBL) also to the problem of generating expressive performances. The IBL approach is used to complement a notelevel rule-based model with some predictive capability at the higher level of musical phrasing. More concretely, the IBL component recognizes performance patterns, of a concert pianist, at the phrase level and learns how to apply them to new pieces by analogy. The approach produced some interesting results but, as the authors recognize, was not very convincing due to the limitation of using an attribute-value representation for the phrases. Such simple representation does not allow taking into account relevant structural information of the piece, both at the sub-phrase level and

at the inter-phrasal level. In a subsequent paper, Tobudic and Widmer [24], succeeded in partly overcoming this limitations by using a relational phrase representation.

The possibility for a computer to play expressively is a fundamental component of the so-called "hyper-instruments". These are instruments designed to augment an instrument sound with such idiosyncratic nuances as to give it human expressiveness and a rich, live sound. To make an hyper-instrument, take a traditional instrument, like for example a cello, and connect it to a computer through electronic sensors in the neck and in the bow, equip also with sensors the hand that holds the bow and program the computer with a system similar to SAXEX that allows to analyze the way the human interprets the piece, based on the score, on musical knowledge and on the readings of the sensors. The results of such analysis allows the hyper-instrument to play an active role altering aspects such as timbre, tone, rhythm and phrasing as well as generating an accompanying voice. In other words, you have got an instrument that can be its own intelligent accompanist. Tod Machover, from MIT's Media Lab, constructed such an hypercello and the great cello player Yo-Yo Ma premiered, playing the hypercello, a piece, composed by Tod Machover, called "Begin Again Again..." at the Tanglewood Festival several years ago.

3 Apparently or really creative?

The main limitation of computational models of creativity is the absence of built-in evaluation criteria to select the most valuable new combinations among the numerous generated ones. This is especially difficult in artistic domains such as music. In these domains the selection is generally done by humans. This is indeed the case of our system SaxEx and of the other systems we know, particularly those based on evolutionary computation. In other domains there have been some attempts to provide the system with evaluation capabilities.

Margaret Boden pointed out that even if an artificially intelligent computer would be as creative as Bach or Einstein, for many it would be just apparently creative but not really creative. I fully agree with Margaret Boden in the two main reasons for such rejection. These reasons are: the lack of intentionality and our reluctance to give a place in our society to artificially intelligent agents. The lack of intentionality is a direct consequence of Searle's Chinese room argument, which states that computer programs can only perform syntactic manipulation of symbols but are unable to give them any semantics. This critic is based on an erroneous concept of what a computer program is. Indeed, a computer program does not only manipulate symbols but also triggers a chain of cause-effect relations inside the computer hardware and this fact is relevant for intentionality since it is generally admitted that intentionality can be explained in terms of causal relations. However, it is also true that existing computer programs lack too many relevant causal connections to exhibit intentionality but perhaps future, possibly anthropomorphic, "embodied" artificial intelligences, that is agents equipped not only with sophisticated software but also with different types of

advanced sensors allowing to interact with the environment, may have enough causal connections to have intentionality.

Regarding social rejection, the reasons why we are so reluctant to accept that non human agents can be creative is that they do not have a natural place in our society of human beings and a decision to accept them would have important social implications. It is therefore much simpler to say that they appear to be intelligent, creative, etc. instead of saying that they are. In a word, it is a moral but not a scientific issue. A third reason for denying creativity to computer programs is that they are not conscious of their accomplishments. However I agree with many AI scientists in thinking that the lack of consciousness is not a fundamental reason to deny the potential for creativity or even the potential for intelligence. After all, computers would not be the first example of unconscious creators, evolution is the first example as Stephen Jay Gould [25] brilliantly points out: "If creation demands a visionary creator, then how does blind evolution manage to build such splendid new things as ourselves?"

4 References

[1] M. Boden,*The Creative Mind: Myths and Mechanisms*, New York; Basic Books 1991.
[2] M. Boden (Ed.); *Dimensions of Creativity*, MIT Press 1994.
[3] T. Dartnall (Ed.); *Artificial Intelligence and Creativity*, Kluwer Academic Pub. 1994.
[4] D. Partridge, and J. Rowe; *Computers and Creativity*, Intellect Books,1994.
[5] P.J. Bentley and D.W: Corne (eds.), *Creative Evolutionary Systems*, Morgan Kaufmann 2002.
[6] J. Rowe, D. Partridge, Creativity: A survey of AI approaches, *Artificial Intelligence Review* 7 (1993), 43-70.
[7] B.G. Buchanan, Creativity at the Metalevel: AAAI-2000 Presidential Address, *AI Magazine* 22:3 (2001), 13-28.
[8] M.L. Johnson. *An expert system for the articulation of Bach fugue melodies*. In *Readings in Computer Generated Music*, ed. D.L. Baggi, 41-51. Los Alamitos, Calif.: IEEE Press. 1992.
[9] R. Bresin. *Articulation rules for automatic music performance*. In Proceedings of the 2001 International Computer Music Conference 2001. San Francisco, Calif.: International Computer Music Association. 2001.
[10] A. Friberg. *A quantitative rule system for musical performance*. PhD dissertation, KTH, Stockholm. 1995.
[11] A. Friberg, R. Bresin, L. Fryden, and J. Sunberg. Musical punctuation on the microlevel: automatic identification and performance of small melodic units. *Journal of New Music Research* 27:3 (1998), 271-292.
[12] A. Friberg, J. Sunberg, and L. Fryden. Music From Motion: Sound Level Envelopes of Tones Expressing Human Locomotion. *Journal on New Music Research*, 29:3 (2000),199-210.
[13] S. Canazza, G. De Poli, A. Roda, and A. Vidolin. *Analysis and synthesis of expressive intention in a clarinet performance*. In Proceedings of the 1997 International Computer Music Conference, 113-120. San Francisco, Calif.: International Computer Music Association. 1997.
[14] R.B. Dannenberg, and I. Derenyi. Combining instrument and performance models for high quality music synthesis, *Journal of New Music Research* 27:3 (1998), 211-238.

[15] R. Bresin. Artificial neural networks based models for automatic performance of musical scores, *Journal of New Music Research* 27:3 (1998), 239-270.
[16] R.A. Kendall, and E.C. Carterette. The communication of musical expression. *Music Perception* 8:2 (1990), 129.
[17] J.L. Arcos, R. Lopez de Mantaras, and X. Serra; Saxex: A Case-Based Reasoning System for Generating Expressive Musical Performances. *Journal of New Music Research* 27:3 (1998), 194-210.
[18] J.L. Arcos, and R. Lopez de Mantaras; An Interactive Case-Based Reasoning Approach for Generating Expressive Music. *Applied Intelligence* 14:1 (2001), 115-129.
[19] R. Lopez de Mantaras, and J.L. Arcos; AI and Music: From Composition to Expressive Performance. *AI Magazine* 23:3 (2002), 43-57.
[20] M. Grachten, J.L. Arcos, and R. Lopez de Mantaras; *TempoExpress, a CBR Approach to Musical Tempo Transformations*. In Proceedings of the 7th European Conference on Case-Based Reasoning (Eds.
P. Funk and P. A. Gonzalez Calero), Lecture Notes in Artificial Intelligence 3155 (2004), 601-615.
[21] P. Desain and H. Honing. *Tempo curves considered harmful*. In "Time in contemporary musical thought" J. D. Kramer (ed.), Contemporary Music Review. 7:2, 1993.
[22] T. Suzuki, T. Tokunaga, and H. Tanaka; *A Case-Based Approach to the Generation of Musical Expression*. In Proceedings of the 16th International Joint Conference on Artificial Intelligence, Morgan Kaufmann (1999), 642-648.
[23] A. Tobudic, and G. Widmer; *Playing Mozart Phrase by Phrase*. In Proceedings of the 5th International Conference on Case-Based Reasoning (Eds. K.D. Ashley and D.G. Bridge), Lecture Notes in Artificial Intelligence 3155 (2003), 552-566.
[24] A. Tobudic, and G. Widmer; *Case-Based Relational Learning of Expressive Phrasing in Classical Music*. In Proceedings of the 7th European Conference on Case-Based Reasoning (Eds. P. Funk and P. A. Gonzalez Calero), Lecture Notes in Artificial Intelligence 3155 (2004), 419-433.
[25] S. J. Gould; Creating the Creators, *Discover Magazine*, October (1996), 42-54.

El (inter)cambio imaginal. IMAGEN: (weid-es: “ver formas”; y gen: “luz”)

Martin Caiero

Universidad de Vigo, Facultad de Bellas Artes

Departamento de Escultura, Campus de Pontevedra

> «El lenguaje nace de la imaginación que suscita y en todo caso excita al sentimiento o a la pasión»
>
> Jean Jacques Rousseau

Nuestra mente todo lo que conoce ha de imaginarlo. «Percibir es ver desde la conciencia, es un desentrañar de las cosas para darles luz, para iluminarlas”. Es lo que media entre la experiencia y el conocimiento; y lo que trasforma lo pensado en inteligible. Pero es preciso distinguir dos cosas: la imaginación como formadora de la realidad individual; y la imagen como forma de existencia social. La imagen en su proceso de creación relaciona a la Ciencia, la Técnica, la Religión, la Política y el Arte. Nuestra cultura experimenta hoy principalmente su existencia a través de imágenes, o como definió Debord: «el espectáculo no es un conjunto de imágenes, sino una relación social mediatizada por imágenes» . La existencia del individuo está inmersa en numerosas imaginaciones que han configurado los cronotopos de su realidad. «De las ciencias puras a las ciencias aplicadas y de éstas a la tecnología el método creativo ha sido aceptado hoy en día por todas las disciplinas científicas, el método por modelización y simulación cambia la condición de la experiencia y de la realidad... la realidad deja paso a lo virtual. » La imaginación se estimula en dos direcciones: una que definió Nietzsche a finales del siglo XIX: «sólo como fenómeno estético puede concebirse la existencia del mundo»; y otra ―como consecuencia de la primera, pues si hay estetización, es porque se realizan operaciones de poetización― de creación. En un régimen imagocrático como el actual, el individuo es intercambiado en forma de imagen. «No hay que manufacturar sólo productos, o mejores productos, sino también imágenes de los productos... La proliferación de imágenes de productos o de empresas que de ello resulta es la que, por un lado obliga a plantearse temáticamente el problema del valor informativo y semántico de las formas, y, por otro, crea un entorno de símbolos o imágenes.» El conocimiento crea, pues el mundo que es capaz de imaginar. Por eso, la reflexión sobre cómo se produce este conocimiento nos dirige inexorablemente al proceso de Creación de imágenes, en las que todo el acto de la percepción se soporta.

A. Fernández-Caballero & S. Miguel (Eds.): 50 Años de la Inteligencia Artificial, pp. 51-65, 2006.

Encontramos pues, en este sentido, que la naturaleza misma posee una actividad muy similar a la imaginación del ser. Cada cosa que existe, tiene la pulsión de devenir en imagen, en imago. Por ejemplo, el espermatozoide y el óvulo intercambian puntadas hasta que el feto se gesta. El bebé se despliega hasta la edad adulta y no deja de trasformarse hasta la vejez. Las mismas operaciones trasformadoras acontecen en todos los reinos, aunque con distintos ritmos imaginales. Este fenómeno nos lleva al ámbito del pro-grama ("lo que está antes de la letra, del gesto"), en el que las existencias se definen. Nuestra época se ha centrado en comprender y controlar el modus operandis de la vida misma en sus orígenes más elementales. El programa genético es imagenético. En los genes se contienen los imagenes. Para que algo exista a la conciencia, debe haberse al menos «dado a luz». Si se trata de saber, y de hacer saber (fabricar, crear artificialmente el conocimiento) deberíamos preguntarnos ¿por qué cambian las cosas, por qué cambian las personas, sus hábitos, sus gustos, sus ropas, sus casas, sus transportes, sus conocimientos, sus fantasías, sus ideas? Cambian por el intercambio. Las cosas cambian su imagen al variar su materia, su energía y su información, o «sus formas de existencia»: su morfogénesis. «Si tomamos un trozo de materia viva como sistema complejo, nadie discute que para que tal materia mantenga su condición de viva se necesita un continuo intercambio de tres magnitudes básicas: materia, energía e información.» . En el intercambio se producen actos que provocan transformaciones en la existencia. Este cambio por el intercambio es lo que he llamado aquí el intercambio imaginal.

Hay en este sentido, dos términos que nos sirven para exteriorizar nuestra situación: Fanerón: «el total de todo lo que de algún modo o en algún sentido está presente en la mente, con independencia absoluta de si corresponde a algo real o no»; y Faneroscopia: equivalente a phenomenon, «todo lo que en cualquier sentido está ante nuestras mentes». Como toda imagen es política y está llena de ideologías, la realidad que el individuo construye siempre exige interpretación y descripción, hermenéutica y heurística. El conocimiento consiste en desvelar e iluminar. La imagen es el soporte en el que se mantiene el conocimiento. Debemos distinguir así que la conciencia del individuo pertenece al ámbito de lo real, que su capacidad para pensar lo distingue de lo natural. Lo que está ahí, lo que se da al individuo abiertamente, todo ello es lo real. Es lo que da lugar a las diversas realidades individuales. Una realidad comparte aspectos con otra realidad acerca de lo real, pero ni en sí mismas ni en su conjunto, esas realidades comportan lo real; son aproximaciones de la conciencia.

Relación de partículas imaginales que conviven en recursividad							
ELEMENTO	**Borde**	**Sociedad**	**Aporte**	**Ámbito**	**Ambiente**	**Ritmo**	**Activación**
Humano	Real	Humanidad	Sujeto	Signo	Biosfera	Biótico	Imaginación
Máquina	Artificial	Maquinidad	Objeto	Código	Tecnosfera	Tecnótico	Ingeniería
Imagen	Virtual	Imagenidad	Símbolo	Tropo	Noosfera	Trópico	Imaginería
Naturaleza	Natural	Naturalidad	Metabolismo	Vacío	Zoosfera	Genético	Imagenería
Hiperbolismo imaginal							

En lo imaginal se mezclan cuatro elementos: Humano, Máquina, Naturaleza, e Imagen. A lo humano le corresponde la inteligibilidad del signo, que provoca la humanidad. A la máquina le corresponde la inteligibilidad del código, que provoca la maquinidad. A la imagen le corresponde el símbolo. A la naturaleza el hueco («vacío»), que provoca la aparición de maquinidad y humanidad. Si existe algún principio activador, de precipitación del intercambio, no es la naturaleza en sí, sino el vacío, el hueco que permite que algo actúe. Como afirman los Kukuya: «la naturaleza del objeto es poco importante, lo esencial es que actúe». Lo imaginal sería de este modo en su conjunto, la articulación de una actividad en la que entran en juego el vacío, el signo, el código y el símbolo. La forma funcional en la que se ponen en acto todos estos agentes de lo imaginal es el tropo (de trep: voltear, girar, cambiar algo de sitio).

La palabra «símbolo» nos recuerda Peirce, «tiene tantos significados que sería infligir un perjuicio al lenguaje el añadirle uno nuevo. Etimológicamente significaba una cosa lanzada unida, al igual que >:%≅&≅< (embolum) es una cosa lanzada dentro de algo, un pestillo, y 0∀Δ□%≅&≅< (parabolum), una cosa lanzada de lado, una garantía subsidiaria, e ∴B &≅&≅< (hypobulum), una cosa lanzada por debajo, un obsequio prenuncial»[i]. En todo sentido de «símbolo», existe una analogía con la célula, un entorno con un núcleo en cuya naturaleza está posibilitado y permitido el intercambio. «Se dice usualmente que en la palabra *símbolo* el lanzar unido hay que entenderlo en el sentido de «conjeturar»... pero los griegos con mucha frecuencia usaban «lanzar unido» (≅β7∃□B&≅<) para significar el hacer un contrato o una convención.»[ii] Además, tenemos el consentimiento de la asociación por la partícula *jet*, presente tanto en ob*je*to como en su*je*to.[iii] En el sentido de símbolo permanece latente el sentido de con*je*tura, que nos hablaría de un contrato. El proceso de fusión simbólica (*simbiólisis*) es lo que provoca *ideas*. Las "ideas" son generadas en la *mente* del individuo. La percepción es el proceso por el que la mente piensa, conoce y actúa[iv]. Percibir es una relación por la que el individuo sale a la intemperie y deja que la experiencia entre en él. Al percibir fluye y es influido. «La mente, nos dice Damasio, existe en un organismo integrado y para él; nuestras mentes no serían como son sino fuera por la interacción de cuerpo y cerebro durante la evolución, durante el desarrollo individual y en el momento presente. La mente tuvo que estar primero relacionada con el cuerpo, o no hubiera existido. Sobre la base de la referencia fundamental que el cuerpo está proporcionado de forma continua, la mente puede estar relacionada después con muchas cosas, *reales o imaginarias*.» (la cursiva es mía)[12] . La mente es capaz de generar ideas, incluso propias, con nuevos tropos, o posibilidades simbólicas que dan lugar a nuevas entidades de código y de signo. Hay cosmogonías personales que son las que permiten al individuo distanciarse y diferenciarse en su realidad, por su "creatividad". Hasta el punto de originar su propia «realidad artificial» (o «realidad real»). Para que una idea se convierta en imagen y pueda hacerse transmisible, deben colaborar máquina y humano; la ingeniería y la imaginación. «Desde un punto de vista cognitivo se diferencia en los procesos mentales, un nivel computacional y otro fenomenológico. El primero engloba los complejísimos sistemas de organización que transforman ciertos estímulos externos en elementos de transmisión entre distintos formatos internos. El segundo engloba los

complejísimos sistemas que transforman esos elementos computacionales en un formato analógico, dando como resultado el mundo fenomenológico de la conciencia y los procesos psíquicos superiores. Estas dos mentes simultáneas y cooperativas, computacional y fenomenológica apuntan a dos tipos de registros: la primera opera con elementos computacionales, en formatos físicos y químicos, electromagnéticos, energéticos, en códigos digitales; la segunda opera con elementos analógicos fabricando modelizaciones de síntesis que acaban convirtiéndose en los contenidos de la conciencia (y del inconsciente). No obstante, el nivel computacional y el nivel fenomenológico comparten la condición de representación, que se define como "sistema de distinciones computables". Pero en cada caso, la modelización remite a parámetros diferentes.» La complejidad se produce cuando lo fenomenológico se une a lo computacional. Esta asociación "biológica" se produce debido a la necesidad humana de generar consciencia de sí misma y del entorno en el que existe. Y esto lo necesita hacer primero para poder percibirse con determinada perspectiva más allá de la imagen .

La sociedad se basa, cuando menos, en principios políticos (praxis: acción), éticos (ethos: comportamiento), poéticos (poiesis: creación), estéticos (aisthesis: percepción) y técnicos (tecné: fabricación). ¿Qué es tan atrayente en la imagen para que numerosos actores y las disciplinas generadas entorno a cada uno de sus saberes acaben influenciados por la actividad imaginal? La imagen transmite ideologías, ideas reflexionadas, trasfiere comportamientos, modos de actuar, de contemplar, procesos de fabricación, de utilización, en definitiva: la imagen HACE culturas. Todo es susceptible de ser programado. En los laboratorios imaginales el estudio y aplicación de programas se hace con ayuda de simuladores. En el ámbito de la imaginación hay dos tipos de programa: el programa genético, portador del gen; y el programa ectogénico, importador o exportador del gen. Un «gen» es lo que «da a luz», lo que hace «parir» a las cosas para que sean imaginales. No presupone la negación de algo anterior a él. El gen porta la información desencadenante del proceso imaginal. El gen es el desencadenante del metabolismo y es lo que activa el proceso recursivo, en el que se activan las informaciones que al desplegarse generan un ser. El programa genético describe el comportamiento del gen desde la fertilización hasta que su despliegue se completa. Es lo que podríamos llamar «una herencia imaginal». Hay ideas génicas y transgénicas. Las primeras las posibilita lo natural y las segundas lo real. Ambas pueden transformar la realidad del individuo, sus pensamientos, sus emociones, sus sentimientos. Las categorías de percepción de lo real cambian y exigen nuevos pensamientos, nuevas formas de sentir, de inteligibilidad, nuevas relaciones de entendimiento y comprensión, en definitiva: nuevas mentalidades . Si los cambios en los niveles fisiológicos, psicológicos, ideológicos transforman «mentalmente el mundo», también se produce el camino inverso: cuando algo cambia en la atmósfera imaginal, se producen cambios físicos, y por lo tanto mentales en la ingeniería individual. El plano ideológico, a través de la fuerza física ha sido capaz de transformar, destruir civilizaciones enteras: los hábitos, los ritos, las costumbres, la gastronomía, sus estamentos políticos, los juegos. La imagen hace culturas, y es capaz de cambiarlas.

Si nos guiamos por la mínima expresión de lo imaginal, el gen, encontramos que lo indígena, viene de la palabra «aborígeno»: "lo que es originario de un lugar". Lo aborigen es lo que no se puede conocer, a lo que no se puede acceder, es, si queremos

entenderlo así, lo que está tachado, borrado de la consciencia. El gen ¿de dónde surge? No lo sabemos. Usamos lo que él nos da, lo que está ahí, su bios (signo) y su tecnos (código), sin poder saber cómo se originó. Sólo sabemos que se genera. Lo «amorfo» es aquello que carece de información, lo que es susceptible de ser informado. La primera situación "aborígena", el estado inicial de toda cultura es la amorfosis. La fórmula imaginal articula y procesa lo analógico y lo anómalo. Ambos son principios de actuación y de activación. La analogía es reversible y mudante, socializa manteniendo los bordes de lo real (con lo virtual y lo artificial) definidos a ambos lados de la relación. Tiene que ver con las formas de la identidad. Pero por sí sola no explica o describe los "cambios" reales que se producen en el intercambio. «Esos procesos analógicos de la mente computacional no son accesibles a la conciencia, tal y como no somos capaces de acceder o intervenir directamente sólo con la conciencia en los procesos de crecimiento: no podemos estar presentes en el procesamiento del que deriva la emisión de tal enzima, o en cada duplicación celular, o en la transformación química de una señal electromagnética que ha excitado ciertas células en la retina.» La anomalía tiene que ver con las formas de la alteridad. Es irreversible y mutante. Explica o describe los "cambios" naturales que se producen en el intercambio.

Consideremos pues, el tratamiento del gen y de la grama como generadores y programadores de existencias, que cualquier forma de existencia articula en una relación bólica (no siempre conocida) genes y gramas. El gen genera y la grama programa; lo imaginal es genética y gramática. La agencia hace su imagenería injertando en el gen real lo que le interesa, su programa específico, en una mimesis asombrosa que repite el modelo por el que la naturaleza actúa durante el intercambio imaginal. «Las palabras eran originariamente imágenes, intuiciones de la realidad debidas al impulso metaforizante del individuo. Ahora bien, ese impulso hacia la construcción de metáforas, ese impulso fundamental del hombre del que no se puede prescindir ni un solo instante, pues si así se hiciese se prescindiría del hombre mismo...» La agencia (de publicidad) ya no hace sólo anuncios, sino programaciones de existencia. Un anuncio es antes que nada un programa preparado para inscribirse como acto imaginal. Este hecho se hace posible gracias que en la atmósfera imaginal, la interactividad entre lo real y lo natural es ejercida por conmovedores: mutadores y mudadores. El «conmutador» y el «conmudador» sirven para establecer el paso de un estado a otro, de lo trópico a lo simbólico, de lo simbólico a lo metabólico. El "conmovedor" pertenece a una «realidad hiperbólica» en la que se permite todo tipo de giros, de cambios. Emociona las cosas haciéndolas salir de sí mismas. Lo Hiperbólico es el hábitat de los signos, los códigos y los vacíos de la naturaleza. El mutador es integral, mientras que el mudador es superficial. Lo conmovedor está así en contacto con lo trópico, es lo que pone en contacto al signo con el código. Los conmovedores impiden, pues la ruptura de los procesos de intercambio, favorecen el tránsito de los tropos de lo metabólico a lo simbólico.

El «tropo» es lo que anima el símbolo (y con él al sujeto y al objeto), es lo que «da sentido a su existencia». Este espíritu es lo que permite actuar a las cosas, y, gracias a ello, lo que posibilita el intercambio por parte de los conmovedores, lo que hace discurso y evita que se acabe en la confusión o en la dispersión parabólica. Esto sólo es posible cuando los discursos están acordados (a pesar de sus trayectorias). «Si todo fuera una diversidad absoluta, el pensamiento estaría destinado a la singularidad,

...estaría destinado a la dispersión absoluta y a la absoluta monotonía. No serían posibles ni la memoria ni la imaginación, ni en consecuencia, la reflexión. Sería imposible comparar las cosas entre sí, de definir sus rasgos idénticos y de fundar un nombre común: no habría lenguaje.» Un solo individuo, una sola parábola no puede comprender la diversidad hiperbólica. La universidad debe comportarse, sensibilizarse con la diversidad y colaborar en la comprensión. «El término universidad, en su sentido latino originario de conjunto universal de todas las formaciones para todo tipo de capacidades» ha sido sobrepasado . Las imágenes, las imaginaciones, lo imaginal... migra. La atmósfera imaginal lo metaboliza todo gracias a su naturaleza hiperbólica. (No es extraño que a los insectos de metamorfosis completa se les denomine como holometábolos, pues actúan motivados por el hiperbolismo).

Máquinas y humanos son vehículos unos de otros porque los conmovedores (mutadores y mudadores) se convierten en vehiculares . Una de las ciencias que más ha estudiado el metabolismo biológico es la termodinámica, «que es, nos dice Wagensberg, una ciencia clave para la comprensión y descripción general del cambio. Su área de mayor impacto: la biología. Porque, por un lado, los sistemas vivos, son los que mayor grado de complejidad muestran en todos los niveles de su estructura.». La biología considera que los seres vivos cambian en base a dos principios: 1) adaptaciones como comportamientos deterministas por "interacción" al medio: repliegues, inflexiones externas.. 2) autoorganizaciones como comportamientos indeterministas, en "interiorización": cambios emocionales, trastornos, reflexiones internas... Para la ciencia, el ser biológico es "un sistema" porque tiene "una estructura". Ningún sistema es posible si no hay una predeterminación estructural sobre el que el sistema se mantenga. La estructura contiene el sistema cuando se utiliza, por ejemplo en el modelo anatómico. Tanto el individuo autónomo como el autómata (o automático) participan de una anatomía.

En la atmósfera imaginal todo es metabólico y susceptible de simbolizarse. Pero los bordes que van de lo macrobiótico a lo microbiótico cambian de naturaleza; sus condiciones espaciotemporales son diferentes (se piensa, por ejemplo, que en el mundo de las partículas existen 6 dimensiones más además de las 4 que encontramos en nuestro mundo corporal). Para imaginar lo imaginal —en todos los momentos del proceso, que va de la partícula, a la molécula, a la célula, al cuerpo o a la cultura —es preciso entonces, cambiar la perspectiva con la que concebimos las cosas, la existencia, a los demás y a nosotros mismos. Y esto es posible porque el cerebro (y el individuo en su conjunto) posee una gran plasticidad, capaz de modelarse a la mentalidad requerida . El paso de una versión a otra es lo que se conoce como la versatilidad de la forma. Ningún ejemplo es tan claro del paso de lo microbiótico a lo macrobiótico como el proceso de reproducción entre humanos. Desde la producción del espermatozoide en las gónadas masculinas, las espermatogonias y la producción del ovocito en las gónadas femeninas, las ovogonias hasta la fecundación del óvulo en el que se gestará el nuevo individuo, se pasará de los cromosomas (nivel celular) a lo corporal y a lo cultural yendo por todas las etapas de la existencia. Desde el parto hasta la partida la programación se despliega despuntando y enhebrando el programa femenino con el programa masculino: la mitad de cada individuo, de su programa genético pertenece a cada uno de los progenitores: un ojo, una mano, un brazo, cinco dedos, un pie, una pierna, un seno, la mitad de cada agujero... El embarazo se produce

porque en el "coito" se suceden los procesos bólicos y versales: el pene y la vagina actúan por embolismo; el esperma y el ovocito se unen en el ovario por simbolismo; el feto se desarrolla recursivamente en el vientre hasta que "rompe aguas" y discurre para empezar su propia parábola, a deambular (pasearse). Un embarazo podría considerarse una forma de embolia, o el brote de las flores: salidas de vainas por inducción genética; desenvainar es ambular. Por eso el programa genético necesita existir en un medio hiperbólico.

En la imaginación coexisten diferentes formas de simbolización: símbolos inmunes, inertes, que generan los residuos de anteriores procesos imaginales; símbolos embólicos, que han alcanzado una estabilidad perfecta; símbolos neo, que desencadenan nuevas formas de existencia. La ruptura de la "conjetura", sin consentimiento, sin "coyuntura", se produce cuando el signo o el código, la máquina o lo humano rompen su relación con el surco en el que estaban; o cuando se abandona un discurso para ir a otro. Se puede producir por coagulación: saturación de flujos formando un coágulo, o por ausencia de sinapsis en los conmovedores. Cuando se produce, en el intercambio con imágenes, más salida de signos que de códigos, o de códigos que de signos acontece una deformación. Son las denominadas aberraciones imaginales.

El individuo biológico en el intercambio imaginal, pues, progresa. Su progresión se interrumpe al encontrarse en su trayectoria con el individuo tecnológico, y con lo virtual. Se produce una especie de rebote existencial en el que la mente pierde su virtualidad (adaptada a la virtualidad de la máquina) que hace re-dimensionar a todo el individuo (en cierto sentido son aberraciones que acontecen en el borde de lo real). La progresión presupone la comprensión, que permite que la individualidad o la colectividad conserven su existencia, aunque con otra realidad. El intercambio imaginal no es una fórmula de "igualdad". La conciencia parece que se ha dirigido hacia el concepto de "complejidad". Tal como reconoce Damasio: «a medida que los organismos adquieren mayor complejidad, las acciones "dictadas por el cerebro" requirieron más procesamiento intermedio. Otras neuronas se interpolaron entre la neurona del estímulo y la neurona de la respuesta, y así se constituyeron variados circuitos paralelos, pero de ahí no se sigue que el organismo con este cerebro más complicado tuviera una mente. El cerebro puede tener muchos pasos intermedios en los circuitos que median entre el estímulo y la respuesta, y seguir careciendo de mente, sino cumple una condición esencial: la capacidad de representar internamente imágenes y de ordenar dichas imágenes en un proceso denominado pensamiento. (Las imágenes no sólo son visuales; también son «imágenes sonoras», «imágenes olfativas», etcétera.). Mi afirmación sobre los organismos con comportamiento puede completarse ahora diciendo que no todos tienen mente, es decir, no todos poseen fenómenos mentales (que es lo mismo que decir que no todos tienen cognición o procesos cognitivos). Algunos organismos poseen a la vez comportamiento y cognición. Algunos tienen acciones inteligentes pero carecen de mente. No parece existir ningún organismo que tenga mente pero no acción. El individuo humano utiliza la conmutación de los símbolos como fuente de sus procesos y retrocesos neurales en los que se basan los fenómenos mentales, por ejemplo la percepción, el aprendizaje, la remembranza, la emoción y el sentimiento, y... el razonamiento y la creatividad. La regulación corporal, la supervivencia y la mente se hallan íntimamente entrelazados.»

Los "imagenes" gobiernan la existencia de los individuos. Influyen en su pensamiento y en su comportamiento y controlan o influencian su actividad imaginativa, la generación de imágenes propias, "imagenes" autógenos (o autoimagenes). «En... El gen egoísta, Dawkins acuña para la historia el término meme y enciende otra polémica. El meme es el gen mental. Así de sencillo. Así de grandioso. El meme es un paquete cultural que se transmite por imitación. Su hábitat natural es el cerebro, allí nace y desde allí coloniza a otros cerebros... Muchos objetarán que el meme no es más que una metáfora... De acuerdo, pero lo mismo ha ocurrido con casi todos los grandes conceptos de la ciencia.» El «meme» es lo que permite que la cultura exista; el gen cultural representa en lo macrobiótico lo que el gen celular en lo microbiótico. Ambos actúan en sus respectivos niveles como formadores de existencia, como conjeturadores de organismos y organizaciones para la vida. Sin los imagenes tan solo existiría una enloquecida dispersión, un caos prebiótico. Con el "gen cultural" se hace posible que aquello que otros imaginan transforme nuestra propia imaginación, ofreciéndonos su conocimiento, algo en lo que pensar, despertando una conciencia común. Comprobamos así, que las imágenes no sólo afectan en un sentido figurado, sino también físico. «Una de las razones de que la noción del yo pueda ser tan frágil, pudiera ser que la mente humana procura sin cesar introducirse en la mente de otras personas. Se ha descubierto que las llamadas "neuronas especulares" remedan las experiencias de otros. Cuando vemos, por ejemplo, que otro recibe un pinchazo doloroso, se nos estimulan neuronas en la región cerebral responsable del dolor. El grupo de Blakemore ha observado que incluso la visión de que a otro lo tocan puede activar neuronas especulares. Mostraron a un grupo de voluntarios vídeos en los que se tocaba a varias personas en uno u otro lado de la cara o del cuello. Los vídeos suscitaron en ciertas áreas cerebrales en los voluntarios las mismas respuestas que se producían cuando eran ellos los tocados en las correspondientes partes de sus cuerpos. La inspiración para este estudio le llegó a a Blakemore cuando conoció a una mujer de 41 años, llamada C. que llevaba esta empatía a extremos sorprendentes: la visión de que alguien era tocado provocaba que C. Sintiera que la tocaban a ella en el mismo lugar de su cuerpo. Esta mujer creía que a todo el mundo le pasaba lo mismo.» Lo que interesa es cómo un determinado discurso (flujo de información) transgrede el borde de lo virtual y activa el organismo en su honda naturaleza, ya que todo flujo de información a nivel escópico es psicosomático, y lo es sin necesidad de ejercer o ejecutar un acto de comunicación. No es necesario que la relación imaginal se produzca "en carne y hueso" para generar especulaciones en los niveles neurales, que llevan al pensamiento y a las emociones a activarse y actuar.

«Teméis el cambio», dice Neo en Matrix. Hasta ese momento, todas las versiones literarias y cinematográficas que trataban de las relaciones entre humano y máquina no habían introducido el mundo de la imagen, el mundo de la imaginación misma. Matrix es «un simulador interactivo neural» que construye mundos virtuales que sustituyen la necesidad de lo real conectado directamente con el sistema nervioso del individuo; ella es su neo-cortex. Al desconectar las conexiones, Morfeo (tenía que ser el que liberase del sueño, el dueño del mundo de los sueños) a través de una pastilla azul (tenía que ser un fármaco, el mundo de la enfermedad el que hiciese despertar a la mente, una especie de psicotrópico) el protagonista, el Señor Anderson es "liberado" de Matrix. Y es "dado a luz". El protagonista, Neo, al despertar tiene

condicionada su imaginación manteniendo «una autoimagen residual, una proyección mental de su yo digital». Este es el famoso diálogo (y discurso) clave de esta película:

Neo: —Entonces ¿esto no es real?

Morfeo:— ¿Qué es real? ¿De qué modo definirías qué es real?: Lo que puedes relacionar con lo que puedes sentir, ver, saborear... Lo real podrían ser señales eléctricas interpretadas por tu cerebro.

Neo, necesitará hacer algo más que desenchufar su cuerpo, su organismo, necesitará "liberar" su mente, su «meme» del programa cultural/neural/ imaginal de Matrix. La mente es la que hace lo que es y no es real. Pero el «cuerpo» no puede vivir sin la mente: ambos son la comunidad del individuo. Todos los «memes» conforman la cultura de una comunidad. Matrix es un programa con armonía de precisión matemática alfanumérica y algorítmica. Pero en todo programa se producen anomalías. En el interior de un programa la anomalía es sistémica y crea fluctuaciones hasta en las ecuaciones más precisas, hace "balancear" las fórmulas —lo que a ciertos niveles se traduce como cambios de comportamiento—. La anomalía puede reconocerse como un virus: una enfermedad con un objetivo: generar más anomalías para romper con la computación. En una máquina perfecta, el Arquitecto, el programador prevé estas fluctuaciones y utiliza las anomalías para perfeccionar su programa, que de este modo se rediseña. En toda programación se generan comunidades de individuos, con funciones predeterminadas y exclusivas: tienen un objetivo y actúan por ese motivo (como las células especializadas). Si no existiese un objetivo, no habría motivación, se perdería su función y el programa fracasaría.

Cuando McLuhan reflexiona acerca de la tecnología, todas las relaciones estudiadas se producen en las extremidades, ninguna se dirigía a los intersticios del cerebro "realmente" tal como se hace ahora. «Todos los artefactos humanos son extensiones del hombre, salidas o expresiones del cuerpo humano o la psique privadas o corporativas. Como expresiones son lenguajes, traslaciones de una forma a otra, ya sea hardware o software: metáforas. Por supuesto que todas las palabras, en cada lenguaje, son metáforas. Desde el punto de vista estructural, una metáfora es una técnica de representar una situación en términos de otra situación. Es decir, en una técnica de conocimiento, de percepción (hemisferio derecho) y no de conceptos (hemisferio izquierdo).» Hoy pensamos que el análisis conceptual es inseparable del análisis perceptual, ya que el primero se construye sobre el segundo, sobre las «posibilidades tangibles del segundo». Al analizar las relaciones en el ámbito de las "extensiones", la multiplicidad de discursos que caracteriza a la posmodernidad favorece la idea de un cerebro complejo (siempre ligado al cuerpo), capaz de hacer frente a la faneroscopia, y de hacerla comprensible.

Recordemos a Neo conectado en Matrix mientras escuchamos las palabras de McLuhan: «Lo que parece surgir como el discernimiento más importante del siglo XXI es que el hombre no fue diseñado para vivir a la velocidad de la luz [del imagen]. Sin el equilibrio de las leyes físicas y naturales, los nuevos medios de comunicación relacionados con el video harán que el hombre implosione sobre sí mismo. Al estar sentado en el cuarto de control de la información, ya sea en el hogar o en el trabajo, recibiendo información a enormes velocidades (de imagen, sonido o táctil) desde todas las áreas del mundo, los resultados podrían ser peligrosamente inflativos y esquizofrénicos. Su cuerpo permanecerá en un solo lugar pero su mente volará hacia el vacío electrónico, estando al mismo tiempo en todos los lugares del

banco de datos... El hombre desencarnado tiene tan poco peso como un astronauta pero puede moverse con mayor velocidad. Pierde su sentido de identidad privada

<table>
<tr><td colspan="8">Situación actual en la que se da prioridad a la imagen culturalmente como valor de(l) cambio</td></tr>
<tr><td colspan="8">Metabolismo:.....Embolismo..... .....:Simbolismo</td></tr>
<tr><td rowspan="5">GEN:</td><td>Naturaleza</td><td>imaginal</td><td rowspan="5">imaginación</td><td rowspan="5">realidad</td><td>Ética</td><td rowspan="2">Culturas</td><td rowspan="5">:IMAGEN</td></tr>
<tr><td>Máquina</td><td rowspan="2">Natural y real</td><td>Estética</td></tr>
<tr><td>Humano</td><td>Poética</td><td rowspan="3">PROGRAMAS</td></tr>
<tr><td colspan="2" rowspan="2">PROGRAMAS</td><td>Política</td></tr>
<tr><td>Técnica</td></tr>
<tr><td>e-nano</td><td colspan="5">TRANSGEN</td><td>giga-nte</td></tr>
<tr><td colspan="8">Exó(p)tico:.....Hipnotismo/Panoptismo.....:Escópico</td></tr>
<tr><td colspan="8">Faneroscopia</td></tr>
</table>

porque las percepciones electrónicas no están relacionadas con ningún lugar. Atrapado en la energía híbrida que despiden las tecnologías de vídeo, estará ante una "realidad" quimérica que abarca todos su sentidos a un grado de distensión, una condición tan adictiva como cualquier droga. La mente, como figura, retrocede hacia el fondo y flota entre el sueño y la fantasía. Los sueños tienen una conexión con el mundo real porque poseen un marco de tiempo y lugar verdaderos (por lo general en tiempo real); la fantasía no posee dicho compromiso.» (la cursiva es mía). El individuo pendula su realidad entre esas dos mentes o cronotopos mentales, el computerizado y el fenoménico. Acontecimiento que convierte la experiencia y existencia en una fantasía dotada de la más absoluta e incierta realidad. Una realidad cambiante y sin sentido común, abierta a la dispersión intempestiva de la sociedad en la que vive, una sociedad "quimérica" por "mágica", enferma de ilusión, alucinada por las fábulas de la Ciencia, el Arte, la Religión, la Política y la Técnica.

Debemos considerar que la imagen gestiona la existencia. En un sistema totalitario, éste gestiona la imagen (o trata de hacerlo) que gestiona la existencia, y con ella a la sociedad misma. Como expresaron Adorno y Horkheimer: «poder y conocimiento son sinónimos». Es preciso romper (embolarse) con la trayectoria comunicativa, y ser imaginales. Vivimos en la SOCIEDAD iMÁGICA.

NOTAS

[1] Juan Carlos Meana, *El espacio entre las cosas*, Diputación de Pontevedra, 2002

[1] Guy Debord, *La sociedad del espectáculo*, Pre-textos, Valencia, 2000

[1] Michel Serres, *Atlas,* Cátedra, Colección teorema, Madrid, 1995, p.233

[1] Xavier Rubert de Ventós, *Las metopías*, Editorial Montesinos, Barcelona, 1984, pp.77-78 Esto se producía entorno a 1984. Desde entonces la imagen ha cambiado, y no basta un modelo semiótico para comprender la atmósfera imaginal.

[1] Jorge Wagensberg, *Ideas sobre la complejidad del mundo*, Tusquets, Barcelona, 2003, p.98

[1] Charles S. Peirce, *El hombre, un signo*, Editorial crítica, Barcelona, 1988

[1] «Dan Sperber ha hablado de un "módulo de metarepresentación" (Spenber, 1994), o de "redescripción representacional" (Karmiloff-Smith, 1992), para referirse a un centro de articulación que permite la asociación y la bisociación de informaciones provenientes de los distintos módulos cerebrales, efectuando no sólo tareas de traducción, de eliminación de redundancia, de jerarquización y unificación, sino también, y de un modo sorprendente, tareas de *transformación, autoorganización y creación*. Este "módulo meta-representacional" funcionaría como "ecualización recursivo" de las modelizaciones modulares, de los distintos campos de representación. Ello incluye sistemas de traducción entre formatos: sólo dentro de las capacidades visuales, se hiperestructuran, se "intertraducen" formatos electromagnético, químico, computacional, modelizaciones en fase de esbozo primitivo, en 2½D, modelos en 3D... (Jackendoff, 1987). Y en esa lógica se intertraducen además formatos correspondientes a cada uno de los módulos cerebrales, y a los tipos de inteligencia, bisociando continuamente lo discursivo (lógico/argumental/verbal: lo letrado y lo cifrado—matemático) y lo no discursivo (visual, mental, praxiológico, sensible: lo iletrado —plasticidad, y lo quasicifrado—iconicidad o cualificado)...» Juan Luis Moraza, *Arte y Saber*. P.52

[1] Peirce, *Op.cit.*, p.156

[1] *Ibi.*. p.156

[1] En francés la partícula *jet* ha seguido manteniendo su escritura *jet-er*, que significa tirar, arrojar, lanzar. La palabra *objeto* significaba etimológicamente «algo lanzado contra». No hay porque dejar de pensar, que *sujeto* podría fácilmente responder en aquellos tiempos, al sentido de «alguien que sostiene lo que se le lanza».

[1] Antonio R. Damasio propone siguiendo la estela de la neurociencia, que «la razón humana depende de varios sistemas cerebrales, que trabajan al unísono a través de muchos niveles de organización neuronal, y no de un único centro cerebral... Por lo que la emoción, el sentimiento y la regulación biológica desempeñan su papel en la razón humana.» *El error de Descartes*, Crítica, Barcelona, 2001, pp.10-11

[1] *Íbid.*, p.13

[1] Juan Luis Moraza, *Arte y Saber*, pp.50-51

[1] El individuo humano está constituido por órganos especializados: ojo, nariz, boca, orejas, manos, pies, corazón, hígado, ano, sexo, uñas, piel... cada órgano generado por un conjunto predeterminado de células especializadas. Genéticamente, estas células recibieron de sus progenitores la información que marcaba un objetivo: producir una parte del individuo, con límites en sus funciones que no le permiten ir más allá de la zona de especialización. Por ejemplo, las células del ojo, nunca pasarán más allá de la zona orbital en la que debe ir el ojo. Podríamos decir que las células (y en ellas los genes) despliegan su objetivo y su información comportándose de manera desorbitada para llegar a alcanzar su objetivo: ocupar el lugar en el futuro individuo, en el imago. Hay una especie de ordenamiento (de mandato) que precede al desorden interno que permitirá y posibilitará a las células actuar de este modo. De todos los órganos, se considera que el cerebro es el más complejo. Según Damasio, «Ahora podemos decir con seguridad que no existen "centros" únicos para la visión, o el lenguaje o, puestos a ello, la razón o el comportamiento social. Existen «sistemas» compuestos por varias unidades cerebrales interconectadas; desde el punto de vista anatómico, pero no del funcional, estas unidades cerebrales no son otras que los

antiguos «centros» de la teoría inspirada frenológicamante; y estos sistemas se dedican en realidad a operaciones relativamente separables, que constituyen la base de las funciones mentales. También es cierto que las distintas unidades cerebrales, en virtud de donde estén colocadas en un sistema, contribuyen con componentes distintos al funcionamiento del sistema, por lo que no son intercambiables. Esto es lo más importante: lo que determina la contribución de una unidad cerebral concreta a la operación del sistema al que pertenece no es sólo la estructura de la unidad, sino también su «lugar» en el sistema.» Se critica así, la extendida idea de los hemisferios derecho e izquierdo con sus funciones centralizadas, aplicándole a uno la capacidad lógico/formal y al otro la de formar imágenes. Como nos recuerda Peirce, toda operación, incluso matemática o lógica implica el uso de diagramas, de grafemas, que sostienen cualquier pensamiento. Hay una materialidad en las operaciones mentales igual que una mentalidad en las operaciones materiales. «Los sentimientos son tan cognitivos como cualquier otra imagen perceptual. En una acción externa nos damos a conocer y en una acción interna nos conocemos. En ambos procedimientos intervienen la emoción (literalmente *movimiento hacia fuera*), y la conmoción que generan un sentimiento o personal o compartido. A través del sentimiento lo humano percibe y experimenta todos los cambios que constituyen la respuesta emocional o conmocional y los convierte en pensamientos.» Las consciencia es así resultado de una emoción y de una conmoción. Es interesante cómo en el lenguaje de Damasio se reconoce la incorporación a los modelos de comprensión de lo humano del lenguaje de las máquinas, en concreto de la informática y de la electrónica. «Para resumir, pues, el cerebro es un supersistema de sistemas. Cada sistema está compuesto por una compleja interconexión de regiones corticales y núcleos subcorticales, todos ellos pequeños pero macroscópicos, que están formados por neuronas, todas las cuales están conectadas mediante sinopsis. (No es raro encontrar los términos «circuito» y «red» utilizados como sinónimos de «sistema». Para evitar confusiones, es importante especificar si se piensa en una escala microscópica o macroscópica. En este texto, a menos que se diga otra cosa, los sistemas son macroscópicos y los circuitos microscópicos.» *Op.cit.* p.44

[1] «El panorama de la investigación científica de vanguardia se ha visto conmocionado por los simuladores. Los científicos conocen bien el valor de un resultado experimental o de un resultado teórico, pero ¿cuál es el valor de un resultado simulado? ¿Es la simulación una especie de experiencia o una especie de teoría? ¿Podemos... eludir las limitaciones al progreso científico impuestas por las limitaciones, cada vez más definitivas de la observación y de la experimentación? Una pregunta alternativa sería: ¿Podemos construir conocimiento prescindiendo del aporte de información del mundo real? No podemos sacrificar el flujo de información, pero sí el que ésta se refiera al mundo real. Podemos, en efecto, inventar otro mundo. Y dejar para más tarde la discusión de su parecido con el real. Dicho de otro modo, aunque la complejidad del mundo real nos impida su observación y experimentación, sí podemos experimentar y observar un mundo simulado. Y para ello disponemos... de una ayuda exosomática a cuyo entusiasta desarrollo asistimos: las máquinas de procesar (velozmente) mucha información, las computadoras.» Jorge Wagensberg, *Ideas sobre la complejidad del mundo*, Tusquets, pp. 95-96 Han pasado 21 años desde estas palabras de Wagensberg, y algunas cosas han cambiado. Las nuevas tecnologías exigen «nuevas biologías» y «nuevas imaginaciones». El avance o innovación

científicos siempre ha ido ligado al desarrollo de sistemas de inscripción de ese conocimiento. La simulación es una nueva *categoría de la percepción*, un nuevo apéndice con el que se accede al mundo. Toda actividad investigadora pertenece y actúa en el ámbito de lo real. Lo que significa que en nuestro pensamiento, nuestra imaginación ve, sabe asociada con la máquina. Percibimos, sabemos también *por máquina.* Nuestro conocimiento se instrumentaliza. Por eso cambia cuando las máquinas cambian. El resultado es siempre de carácter imaginal. Todo conocimiento (científico, religioso, filosófico, artístico) es, ya lo hemos dicho, altamente imaginativo, tiene dosis de realidad y de fantasía. Existe algo «mágico» o, por contribuir a nuestro pensamiento, *imágico* en los mundos natural, humano y máquina que favorecen las imágenes.

[1] De ahí los dilemas acerca de la clonación, el uso de trasgénicos, de células madre, de la experimentación con animales con supuestos "fines terapéuticos", etc...

[1] *Íbid.*p.52

[1] Jacques Derrida, *De la Gramatología*, Siglo XXI, Mexico, 2003

[1] Michel Foucault, *Las palabras y las cosas,* Siglo XXI, Madrid, 2005, p.124

[1] El carácter «polinizador» de la imaginación—y de la imagen— ha pervertido el término "disciplinar". Hoy se usan con igual sentido los de "multidisciplinar" e "interdisciplinar" para referirse a la transversalidad de las disciplinas. Lo multidisciplinar sería un primer corrimiento en el interior de un determinado saber (parabolismo), por ejemplo en Bellas Artes, donde las antiguas disciplinas de Dibujo, Escultura, Pintura, Grabado, Vídeo, Fotografía, Diseño... se acercan dejándose intervenir procesual y creativamente. Lo interdisciplinar sería un segundo paso en el proceso donde esa «multidisciplinariedad» se llevaría a otros saberes (embolismo), por ejemplo, a la antropología, a la filosofía, a la política, a la ingeniería, el mercado, etc. Como afirmó Foucault, «las disciplinas disciplinan la mente», y si la mente es el lugar en el que la imagen adquiere sentido y favorece la «consciencia de sí mismo», el esquema de comprensión cognitivo debe estar a su vez abierto a la transformación de los intercambios (deambular). De tal modo que ya no podemos saber si es lo cognitivo o el medio imaginal los que provocan el cambio. En todo caso, se ha sugerido la opción del término *indisciplinar*, donde la mente no esté disciplinada. ¿Pero, qué sociedad podría admitir una materia de ese tipo, de ideal de estudio? La indisciplina es, aparentemente, lo más opuesto al orden, o si queremos, a la institución de un determinado orden; ofrece una idea de torbellino, de disturbio. Devenir indisciplinado es la primera acción que nos libera de la confusión que genera poseer un sistema de percepción distinto al entorno. El gesto indisciplinar es entonces una especie de *clinamen* que nos hace «cambiar de estado». Tanto nos da si el origen de esa inquietud nos llega del exterior o sale del interior. En ambos casos somos nosotros quienes sufrimos un acto «tropológico». Por eso la decisión de salir de los límites de una disciplina exige asimilar que estamos abandonando un determinado "universo cognitivo" para entrar en una abertura llena de corrientes e influencias. Según Simón Marchan Fiz, el uso del término *end-disciplinar*, donde se acabe con las disciplinas es lo que la conciencia desea, el *fin-disciplinar*. O lo que es lo mismo: un conocimiento nuevo, que ya no puede tender a lo universal: *diversificación...*

[1] Lo que motiva que una máquina cambie (de forma, de anatomía, de composición, de metabolismo) es el agotamiento imaginal. Que su programa simbólico pierda energía, se gaste por el uso o por el gusto. ¿Por qué cambia un vehículo? El medio

que pone en circulación el nuevo diseño es el Anuncio, y la disciplina que lo transforma de herramienta para desplazarse, de maquina en símbolo, en un ser trópico es *la publicidad.* Esta disciplina se especializa en anunciar el nuevo vehículo, pero no sólo eso, sino que lo hace por un proceso de enunciación y enseñanza, a través de in-signar con numerosos tropos la máquina al público. Para lograr el necesario impacto visual, la *a-gencia* publicitaria manipula los imagenes de la máquina empleando las más variopintas estrategias de imagenería, estableciendo relaciones y situaciones. La semiótica a analizado pormenorizadamente el funcionamiento de "este lenguaje" cuyos mensajes son las imágenes, cargadas de funciones que exceden al, digamos «gen tópico» de la máquina (que es facilitar la conducción y el desplazamiento). Se injertan en el programa numerosos discursos. Convertido en imagen, trasgredido en sus condiciones reales, el vehículo es conmovedor de todo lo que se relacione imaginalmente con él. Como nos dice el semiótico Umberto Eco, «Un automóvil puede ser considerado desde diversos niveles (desde diversos puntos de vista): a) *nivel físico* (tiene un peso, está hecho de metal y de otros materiales); b) *nivel mecánico* (funciona y cumple una función determinada con arreglo a ciertas leyes); c) *nivel económico* (tiene un valor de cambio, un precio determinado); d) *nivel social* (tienen cierto valor de uso, a la vez que indica ciertos status); e) *nivel semántico* (se injerta en un sistema de unidades semánticas con el que guarda algunas relaciones estudiadas por la semántica estructural, relaciones que siempre son las mismas aunque cambian las formas significantes con las cuales las indicamos; es decir, aunque en vez de /automóvil/ digamos /car/ o /coche/.» Umberto Eco, *La estructura ausente,* Lumen, Barcelona, 1994 Todos estos niveles interactúan de tal forma que sus bordes son indiscernibles en la imagenería publicitaria. Su programa génico se complementa con injertos trópicos. Como conmovedor, el automóvil (y por extensión las máquinas) absorben propiedades como «velocidad», «seguridad», «comodidad», «riqueza», y estas propiedades pueden servir para alterar e influir no sólo en su diseño, sino en otras formas de existencia que se benefician de lo que un determinado anuncio «conmueve». El vehículo es un *intercambiador*, y un símbolo (máquina) a escala real que porta en su interior lo humano.

[1] «La idea cartesiana de una mente separada del cuerpo bien pudo haber sido el origen, a mediados del siglo XX, de la metáfora de la mente como un programa informático». Damasio, *Op.cit.* Hoy pensamos que la mente es polidimensional, que no se genera en un espacio virtual de dos dimensiones. Por eso el estudio del cerebro articula diferentes campos: para una *neuroimaginería* que satisfaga la idea de una mente compleja es necesario investigar transversalmente desde *la neurobiología, la neuroanatomía, la neurofisiología y la neuroquímica*. Damasio parece coincidir en su idea de un «atlas de la mente humana» con la «imagen» de Serres acerca del Arlequín: «Todo Atlas, muestra modelos espaciotemporales de la diversidad en mosaico, imagen final del lugar, del tiempo y de redes heterogéneos, reino animal y vegetal... de Arlequín,... estancias diferentes... provista de sus pasillos... El lugar se viste... con la capa del Arlequín.» (*Atlas,* pp.57-58) Y Damasio: «Una representación de la piel podría ser el medio natural de significar el límite corporal porque es una interfase dirigida a la vez hacia el interior del organismo y hacia el ambiente con el que el organismo interactúa... Este mapa dinámico del organismo anclado en un esquema corporal y un límite del cuerpo no se conseguiría en una única área cerebral, sino en varias áreas, mediante pautas coordinadas de actividad neural... En otras

palabras, el comportamiento dinámico de mapas en que pienso es «somatomotor» (pp.214-215 Por otra parte, el cronotopo de la conciencia, en neurociencia se identifica como neocorteza, el *neocortex*, que es tejido nuevo, «la parte evolutivamente moderna de la corteza cerebral».

[1] Damasio, *Op.cit.*, p.121

[1] Wagensberg, *Op.cit.*pp.77-78

[1] *La neurobiología del yo*, Investigación y Ciencia, Enero 2006

[1]Marshall McLuhan-B.R. Powers, *La aldea global*, Gedisa, colección "El mamífero parlante", Barcelona, 2002, p.43

[1] *Íbid,* p.10

---

[i] Peirce, *Op.cit.*, p.156

[ii] *Ibi.*. p.156

[iii] En francés la partícula *jet* ha seguido manteniendo su escritura *jet-er*, que significa tirar, arrojar, lanzar. La palabra *objeto* significaba etimológicamente «algo lanzado contra». No hay porque dejar de pensar, que *sujeto* podría fácilmente responder en aquellos tiempos, al sentido de «alguien que sostiene lo que se le lanza».

[iv] Antonio R. Damasio propone siguiendo la estela de la neurociencia, que «la razón humana depende de varios sistemas cerebrales, que trabajan al unísono a través de muchos niveles de organización neuronal, y no de un único centro cerebral... Por lo que la emoción, el sentimiento y la regulación biológica desempeñan su papel en la razón humana.» *El error de Descartes*, Crítica, Barcelona, 2001, pp.10-11

Razonamiento Formal

María Manzano

Filosofía
Campus Unamuno. Edificio FES
37007 Salamanca
mara@usal.es

1. Algunos razonamientos

1.1. Novelas y cuentos

En la literatura la imagen de la lógica se asocia con la argumentación, la deducción y el silogismo. Aunque no es éste su único cometido ya que también está ligada a la computación y a las propias máquinas que la llevan a cabo, a la reflexión sobre todo el proceso, a la gestión y transmisión del conocimiento y de la información, a la fundamentación de la matemática así como de otras disciplinas, a la teoría de juegos y de la acción, a la interacción persona-ordenador, a la web semántica, etc.

Example 1. **La conquista del aire**[1]

"¿Estás proponiendo que guardemos el dinero en casa?", preguntó ella, y luego, sin darle tiempo a contestar, le llamó puritano. Entonces Carlos la abrazó rogándole que lo olvidara. Porque en la palabra "puritano" se condensaba un argumento que él ya conocía: "Para querer hay que mancharse. Los puritanos no se manchan. Luego, tú no me quieres". Era lo que Carlos llamaba el Silogismo del reproche.

También en los chistes se emplea la simplicidad aparente de la silogística para traer por los pelos resultados inverosímiles. De esta guisa es el *"Razonamientos tontos y corolarios chungos"* que recibimos por correo electrónico de vez en cuando.

Example 2. **Razonamientos tontos**

Razonamiento 1 Dios es amor. El amor es ciego. Stevie Wonder es ciego. *Luego*, Stevie Wonder es Dios.

Razonamiento 2 Me dijeron que no soy nadie. Nadie es perfecto. Luego, yo soy perfecto. Pero sólo Dios es perfecto. Por tanto, yo soy Dios. Si Stevie Wonder es Dios, yo soy Stevie Wonder.

Corolario chungo *¡¡¡¡ Por Dios, soy ciego!!!!*

Incluso en los cuentos infantiles los personajes extraen conclusiones lógicas de la información de que disponen.

Example 3. MOUSE SOUP[2]

[1] Belén Gopegui. [1998]. *La conquista del aire.* Anagrama.
[2] Arnold Lobel. [1977]. *Mouse Soup.* HarperCollins.

A. Fernández-Caballero & S. Miguel (Eds.): 50 Años de la Inteligencia Artificial, pp. 67-89, 2006.

A mouse
sat under a tree.
He was reading a book.
A weasel
jumped out
and caught the mouse.
"Ah" said the weasel.
"I am going to make
mouse soup."
"Oh" said the mouse.
"I am going to *be*
mouse soup".

Y por supuesto, nuestro paradigmático Mr Sherlock Holmes que llega a sorprendentes conclusiones en el espacio de un segundo.

Example 4. **Estudio en Escarlata**[3].

—Doctor Watson, míster Sherlock Holmes —anunció Stamford a modo de presentación.
—....Por lo que veo, ha estado usted en tierras afganas.
...
—Alguien se lo dijo, sin duda.
—En absoluto. Me constaba esa procedencia suya de Afganistán**.** ... me vi abocado a la conclusión... Helos aqui puestos en orden. "Hay delante de mí un individuo con aspecto de médico y militar a un tiempo. Luego se trata de un médico militar. Acaba de llegar del trópico, porque la tez de su cara es oscura y ése no es el color suyo natural, como se ve por la piel de sus muñecas. Según lo pregona su macilento rostro, ha experimentado sufrimientos y enfermedades. Le han herido en el brazo izquierdo. Lo mantiene rígido y de forma forzada... ¿en qué lugar del trópico es posible que haya sufrido un médico militar semejantes contrariedades, recibiendo además una herida en un brazo? Evidentemente, en Afaganistán"

1.2. *¿Qué había antes?*

La filosofía, la cosmología e incluso la religión se preguntan

¿Qué había antes?

y razonan para llegar a una respuesta. En *La historia más bella del mundo* el periodista Dominique Simonnet entrevista, entre otros, al astrofísico Hebert Reeves y producen un relato interesante del origen del mundo en el que no faltan razonamientos. Veamos un par de ejemplos:

[3] Arthur Conan Doyle. 1887. *Estudio en Escarlata.* (traducción en Alianza Editorial).

Example 5. **La oscuridad de la noche: Una prueba de la Teoría del Big Bang** [4]

El gran descubrimiento de este siglo es que el universo no es inmóvil ni eterno, como supuso la mayoría de los científicos del pasado. ...el universo tiene una historia, no ha cesado de evolucionar, enrareciéndose, enfriándose, estructurándose. ...esta evolución sucede desde un pasado distante que se sitúa, según las estimaciones, hace diez o quince mil millones de años...(cuando) el universo está completamente desorganizado, no posee galaxias, ni estrellas, ni moléculas, ni tan siquiera núcleos de átomos... Es lo que se ha llamado el BIG BANG". Una de las pruebas indirectas de esta teoría se puede plantear así: "Si las estrellas fueran eternas y no cambiaran nunca, como pretendía Aristóteles, la cantidad de luz emitida sería infinita. El cielo debería ser, entonces, extremadamente luminoso. ¿Por qué no lo es? Este enigma atormentó a los astrónomos durante siglos. Ahora sabemos que el cielo es oscuro porque la estrellas no existieron siempre.

En la tradición filosófica que perduró dos milenios se consideraba, como Aristóteles, que el universo era eterno y no cambiaba. Hoy se sabe que las estrellas nacen y mueren tras vivir millones de años. Algunos filósofos lo supusieron, como Lucrecio, en el siglo I antes de Cristo.

Example 6. **Lucrecio, filósofo romano**

Lucrecio afirmaba que el universo aún estaba en su juventud. Razonó así: He comprobado desde mi infancia, se dijo, que las técnicas se han ido perfeccionando. Han mejorado el velamen de nuestros barcos, inventado armas más y más eficaces, fabricado instrumentos musicales más refinados... ¡Si el universo fuera eterno, todos estos progresos habrían tenido tiempo de realizarse cien, mil, un millón de veces!

El origen de los tiempos preocupó también en la filosofía Zen.

Example 7. **El jade celeste.**

Tang de Ying preguntó a Ge: "¿Existían las cosas al principio de los tiempos?" Xia Ge respondió: "Si al principio de los tiempos no hubiesen existido las cosas, ¿cómo sería posible que existiesen hoy? Con idéntica razón, los hombres del futuro podrían decir que hoy no existían las cosas".

Los problemas de cosmología son también el argumento de este divertido pasaje:

Example 8. Confucio se dirigía hacia el este cuando vio a dos chiquillos discutiendo. Al preguntarles el motivo de su disputa uno de ellos dijo: "Yo digo que cuando el sol sale está cerca y a mediodía, lejos. Éste dice que cuando sale está lejos y a mediodía, cerca." Uno argumentaba: "Cuando el sol sale es grande,

[4] Este ejemplo está sacado del libro: *La historia más bella del mundo.* Hubert Reeves y otros. Anagrama: 1997. (páginas 20 y 33)

como un toldo de carruaje. A mediodía, en cambio, del tamaño de un plato o un tazón. ¿Y no es cierto que lo grande está cerca y lo pequeño lejos?" El otro: "Cuando el sol sale es frío y a mediodía, como agua hirviendo. Y lo caliente está cerca y lo frío lejos, ¿no es así?" Confucio no supo resolver el problema. Los dos chiquillos se echaron a reir: "¿Quién dice que tú eres un hombre de grandes conocimientos?"

2. Lógica formal

Los anteriores no son más que una pequeña colección de ejemplos en donde aparentemente se usa la lógica, aunque de manera informal. Todos nosotros, supuestos seres racionales, empleamos la lógica cuando razonamos, asimilamos o procesamos la información que recibimos del entorno, cualquier tipo de información: somos lógicos porque somos seres humanos, y el comportamiento racional implica usarla como herramienta.

Pero la *Lógica* es fundamentalmente una disciplina en sí misma, una de las grandes ramas del conocimiento, que se define como el estudio de la *consecuencia* —esto es, la que se ocupa de los razonamientos válidos o correctos— o, de forma equivalente, como el estudio de la *consistencia* —a saber, la que puede identificar a los conjuntos de creencias compatibles, coherentes, *consistentes, satisfacibles*—

En sentido coloquial se usa el adjetivo lógico no sólo para describir las reglas del razonamiento correcto, sino en una gran variedad de casos, más en concordancia con el uso original del *"logos"* de los griegos, relacionándolo con el lenguaje, la doctrina, la estructura del conocimiento, la razón, etc. Por supuesto, también está directamente emparentada con las matemáticas, que cuentan desde la antigüedad con un ejemplo paradigmático de proceder lógico, *Los Elementos* de Euclides. La Geometría quedó así fijada en unos axiomas de los que mediante reglas de deducción se extraían todos los teoremas generales que la constituían. Es importante señalar aquí la importancia de poder determinar si ese conjunto de axiomas es consistente y también la de poder contar con un cálculo en el que se puedan demostrar todas sus consecuencias, pero sólo ellas. Éste y otros objetivos similares pero más exigentes constituyen el denominado *programa de Hilbert*. Su idea era explotar al máximo la naturaleza finita de las pruebas para proporcionar una fundamentación de la matemática. Podría resumirse su concepción diciendo que preconizaba una axiomatización de las teorías matemáticas de la que pudiera probarse su:

1. *Consistencia.* Es decir, que nunca se podrá demostrar como teoremas de la teoría una sentencia y su negación
2. *Completud.* Es decir, que cada sentencia —del lenguaje en el que se axiomatizó la teoría— sea ella misma o su negación un teorema de la teoría axiomática
3. *Decidibilidad.* Es decir, que exista un procedimiento efectivo o algoritmo mediante el cual, en un número finito de pasos, se determine si una sentencia del lenguaje es o no un teorema de la teoría

Los sistemas de cálculo de Gentzen condujeron a la teoría de la demostración por sus actuales derroteros, ligada inexorablemente a la perspectiva informática. El teorema de Herbrand de 1930 y, posteriormente, el de Robinson se consideran los pilares de la *demostración automática de teoremas*. Hoy sabemos que como programa general el de Hilbert es inaplicable ya que hay resultados negativos que sitúan la capacidad de los formalismos en unas metas menos ambiciosas.

Johan van Benthem [9] entiende que aunque la lógica no es ya el puerto seguro frente a las tempestades del océano de las contradicciones, si que puede considerarse el *sistema inmunológico y dinámico de la mente*.

Durante el transcurso del siglo XX la lógica fue retomando su extensión y amplitud originales estudiándose en ella no sólo el razonamiento matemático sino también fenómenos de gestión y transmisión de información, de toma de decisiones y de la acción, y en general en casi todos los contextos gobernados por reglas. Más importante aún, la lógica constituye el sustrato teórico de la computación, la clave codificadora de sus circuitos internos; pero también se pregunta por su alcance y sus límites. Saber qué pueden y que no pueden hacer los algoritmos, los cálculos deductivos, los lenguajes formales, es crucial y la lógica también se ocupa de ello. El campo de la lógica no se agota en el cálculo que un humano o una máquina pueda efectuar ya que también le interesan las interacciones entre los agentes que participan en la conversación, el proceso de adquisición de conocimiento, la dinámica y el flujo de la información. Otro aspecto a tener en cuenta, pues lo realizamos continuamente en nuestra vida, es el de modificar y revisar nuestras creencias. Por su incidencia en el proceso de adquisición de conocimiento y de mantenimiento de la consistencia de nuestra base de conocimientos lo debemos incorporar a nuestros programas informáticos, a nuestros sistemas expertos. La lógica es argumentación y como a ella se la puede considerar como un juego en el que los participantes emplean ciertas estrategias y sus movimientos están determinados no sólo por sus propios objetivos sino también por los movimientos de sus oponentes.

3. Para empezar: lógica clásica

Hacer lógica formal a partir de un planteamiento intuitivo e informal significa ir soltando lastre. Se eliminan los enunciados del castellano introduciendo un lenguaje riguroso, después, el concepto intuitivo de consistencia como compatibilidad de enunciados, que hace referencia a situaciones posibles, se sustituye por el de satisfacibilidad, en donde las situaciones se reemplazan por las interpretaciones, matemáticamente definidas. A continuación se abandona el concepto intuitivo de consecuencia y se define matemáticamente, en términos semánticos. Así la *lógica* se convierte en una disciplina rigurosa, formal. Más adelante incluso se supera el concepto de validez, ligado al de interpretaciones o modelos, y se introduce la noción de cálculo deductivo como manipulación meramente sintáctica de las fórmulas del lenguaje formal. El objetivo es que al concepto intuitivo le correspondan uno semántico y otro sintáctico, siendo estos últimos equivalentes.

3.1. Lógica proposicional

Reformulemos levemente el argumento del ejemplo 5 para explicitar sus extremos; eliminemos la información supérflua, las preguntas retóricas. Emplearemos un lenguaje muy simple, proposicional, con letras minúsculas —p, q, r, etc.— para representar enunciados atómicos y los signos

$$\neg \qquad \wedge \qquad \vee \qquad \rightarrow \qquad \leftrightarrow$$

para conectarlos entre sí. La formalización que propongo de la prueba indirecta del Big Bang es:

- $(p \rightarrow q) :=$ *Si las estrellas fueran eternas, entonces la cantidad de luz emitida sería infinita.*
- $(q \rightarrow r) :=$ *Si la cantidad de luz emitida fuera infinita, entonces el cielo debería ser extremadamente luminoso.*
- $\neg r :=$ *El cielo es oscuro.*

LUEGO,

- $\neg p :=$ *Las estrellas no existieron siempre.*

Para expresar que la última es una consecuencia de las otras tres escribimos:

$$\{(p \rightarrow q), (q \rightarrow r), \neg r\} \models \neg p$$

Claramente el esquema argumental no levanta sospechas, otra cosa es si aceptáis como verdaderas en el mundo real las hipótesis. Obviamente, el determinarlo no es misión de la lógica. En el presente ejemplo lo sería de la Cosmología. Pero, si el esquema anterior correspondiese a un razonamiento correcto, lo seguiría haciendo cuando retrotradujésemos al castellano p, q y r. De hecho, el razonamiento del ejemplo 6 sigue exactamente el mismo patrón. Por consiguiente, será también correcto: *un razonamiento correcto nos da la pauta de muchos otros*.

¿Cómo se demuestra que es correcto?

En un cálculo axiomático o de deducción natural emplearíamos la regla denominada Modus Tollens —de $\alpha \rightarrow \beta$ y $\neg\beta$ se sigue $\neg\alpha$—, que junto a la de Modus Ponens —de $\alpha \rightarrow \beta$ y α se sigue β— se remontan a la antigüedad clásica. Cuando, como aquí haremos, la consecuencia se prueba sintácticamente en un cálculo deductivo empleamos el signo $\vdash$. Concretamente probaremos

$$\{(p \rightarrow q), (q \rightarrow r), \neg r\} \vdash \neg p$$

En una demostración formal en un cálculo axiomático procederíamos así:

1	$(p \rightarrow q)$	premisa
2	$(q \rightarrow r)$	premisa
3	$\neg r$	premisa
4	$\neg q$	modus tollens 2,3
5	$\neg p$	modus tollens 1,4

Chequear si una demostración formal es correcta es relativamente fácil, basta con comprobar la adecuada aplicación de las reglas del tipo de las mencionadas: Modus Ponens y Modus Tollens. Demostrar en un cálculo puede resultar tedioso: un demostrador automático de teoremas usa millones de pasos similares. Encontrar soluciones es algo más complejo aunque en el caso proposicional hay un método que lo resuelve[5].

3.2. Limitaciones de la lógica proposicional

Pese al buen comportamiento de su cálculo deductivo, al ser la capacidad expresiva de la lógica proposicional extraordinariamente limitada, no nos resulta útil en muchos casos.

Considerad el razonamiento del ejemplo 1, convenientemente reformulado:

- $A :=$ *Carlos es un puritano*
- $B :=$ *Para querer hay que mancharse*
- $C :=$ *Los puritanos no se manchan*

LUEGO:

- $D :=$ *Carlos no ama a Ana*

En lógica proposicional A, B, C y D se formalizan como letras proposicionales —por ejemplo, p, q, r y s— y por lo tanto $\{p, q, r\} \nvDash s$. Sin embargo, el razonamiento es claramente correcto. En primer orden será fácil demostrar la validez del razonamiento.

3.3. El lenguaje de primer orden

Se añade al proposicional la capacidad de analizar las fórmulas atómicas mediante relatores, functores y constantes y la cuantificación e igualdad sobre individuos. Nuestro *Silogismo del reproche* se podría formalizar así:

$$A := Pc \qquad B := \forall x(\exists y Axy \rightarrow Mx) \qquad C := \forall x(Px \rightarrow \neg Mx)$$
$$\textit{Conclusión:} \qquad D := \neg Aca$$

Nuestro objetivo es demostrar que

$$\{Pc,\ \forall x(\exists y Axy \rightarrow Mx),\ \forall x(Px \rightarrow \neg Mx)\} \models \neg Aca$$

En este caso hemos tenido que decidir qué lenguaje formal íbamos a emplear, prescindir de los deícticos y dar nombre a los protagonistas para así eliminar la ambigüedad de la frase *"tú no me quieres"*. La formalización ya no es tan trivial como en el ejemplo anterior, aunque también es sencilla. Concretamente, introducimos las constantes individuales a y c para nombrar a los protagonistas, los relatores M, P y A para *"mancharse"*, *"ser puritano"* y *"amar a"*.

[5] Lo veremos en la sección 6.2.

Para interpretar las fórmulas del lenguaje formal empleamos estructuras matemáticas. En este caso una estructura adecuada

$$\mathcal{A} = \langle \mathcal{U},\ a^{\mathcal{U}},\ c^{\mathcal{U}},\ M^{\mathcal{U}},\ P^{\mathcal{U}},\ A^{\mathcal{U}} \rangle$$

consta de un universo $\mathcal{U}$ no vacío, dos individuos destacados $a^{\mathcal{U}},\ c^{\mathcal{U}} \in \mathcal{U}$, dos subconjuntos del universo $M^{\mathcal{U}},\ P^{\mathcal{U}} \subseteq \mathcal{U}$ y una relación binaria $A^{\mathcal{U}} \subseteq \mathcal{U} \times \mathcal{U}$. En esta estructura A será verdadera (y escribimos $\mathcal{A} \Vdash A$) syss $c^{\mathcal{U}} \in P^{\mathcal{U}}$. Mientras que $\mathcal{A} \Vdash C$ syss $P^{\mathcal{U}} \subseteq\sim M^{\mathcal{U}}$; esto es, el conjunto $P^{\mathcal{U}}$ es un subconjunto del complementario de $M^{\mathcal{U}}$. Finalmente $\mathcal{A} \Vdash B$ syss $Dom(A^{\mathcal{U}}) \subseteq M^{\mathcal{U}}$. Dada una estructura $\mathcal{A}$ no es difícil comprobar que siempre que $A,\ B$ y C son verdaderas tambien lo es D. Pero para demostrar que efectivamente $\{A, B, C\} \models D$ no basta con comprobarlo en una muestra finita de modelos, habría que hacerlo en todos y esto es imposible. Afortunadamente, como en el caso proposicional, hay una forma mucho más sencilla de comprobar consecuencia que no pasa por chequear la verdad de nuestras fórmulas en todo modelo posible: podemos usar un cálculo deductivo similar al empleado en el ejemplo 6 y derivar en él D usando $\{A, B, C\}$ como hipótesis. Recordad que para consecuencia sintáctica (deducibilidad) escribimos $\{A, B, C\} \vdash D$. Por supuesto, el concepto semántico de consecuencia y el sintáctico de deducibilidad se corresponden, lo expresan los conocidos teoremas de completud y corrección, cuya demostración es imprescindible para poder confiar en un cálculo deductivo.

Usamos la lógica como lenguaje en el que representar el conocimiento e interpretamos sus fórmulas en estructuras matemáticas, pero también el cálculo lógico sirve para determinar si el razonamiento de la protagonista es correcto. Para demostrarlo procederíamos así:

1 Pc	premisa
2 $\forall x(\exists y Axy \rightarrow Mx)$	premisa
3 $\forall x(Px \rightarrow \neg Mx)$	premisa
4 $Pc \rightarrow \neg Mc$	eliminación generalizador, 3
5 $\neg Mc$	modus ponens 1 y 4
6 $\exists y Acy \rightarrow Mc$	eliminación generalizador, 3
7 $\neg \exists y Acy$	modus tollens 5, 6
8 $\forall y \neg Acy$	regla derivada
9 $\neg Acp$	eliminación generalizador, 8

Con el paso del tiempo los sistemas de cálculo se han ido perfeccionando y también se han creado otros que son más fácilmente implementables; entre ellos destacan los de tableaux semánticos y los de resolución. Estos cálculos tienen una inspiración claramente semántica ya que sus reglas explicitan las condiciones de verdad de las conectivas y cuantificadores y nos ayudan a encontrar realizaciones o modelos de nuestros enunciados; en el caso proposicional hay básicamente dos reglas

- $\underline{\alpha\text{-reglas } (\alpha = \alpha_1 \wedge \alpha_2)}$: Conjuntivas

y

- $\underline{\beta\text{-reglas } (\beta = \beta_1 \vee \beta_2)}$: Disyuntivas

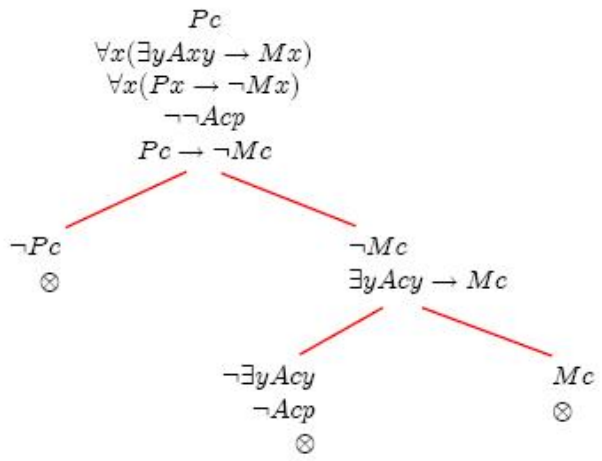
Las β abren dos ramas, una para β_1 y otra para β_2, y las α obligan a que se den ambas, α_1 y α_2, en la misma rama. Hay reglas γ y δ para la cuantificación.

$\gamma-$reglas Si t es un término cerrado
1. de $\forall x A$ podemos deducir $A(t)$
2. de $\neg\exists x A$ podemos deducir $\neg A(t)$

$\delta-$reglas: Si c es una constante que no haya sido usada aún en la rama
1. De $\exists x A$ podemos deducir $A(c)$
2. De $\neg\forall x A$ podemos deducir $\neg A(c)$

Las ramas se van cerrando al encontrarse contradicciones explícitas. Un tableau con todas las ramas cerradas muestra la imposibilidad de encontar un modelo y por consiguiente, cuando queremos demostrar que una fórmula es consecuencia de un conjunto de fórmulas lo que hacemos es seguir un procedimiento refutativo, viendo la imposibilidad de que se den a la vez la hipótesis y la negación de la conclusión: demostramos que $\Gamma \cup \{\neg\varphi\}$ carece de modelo. De hecho demostramos $\Gamma \vdash_{tab} \varphi$ porque también para el cálculo de tableaux de primer orden se aplican los teoremas de completud y corrección mencionados anteriormente.

Concretamente, en nuestro ejemplo el tableau quedaría así



En el ejemplo 3 de la sopa de ratón podemos emplear un lenguaje de primer orden con un relator monario, dos binarios y tres constantes individuales.

$Rx :=$ x es un ratón
$Cxy :=$ x cocina y
$Ixy :=$ x es un ingrediente de y
$w :=$ comadreja del cuento
$m :=$ ratón del cuento
$s :=$ sopa de ratón del cuento

El argumento que le llevó a la conclusión de que iba a ser sopa de ratón

$$D := Ims$$

parte de la hipotesis de que la comadreja declara que va a cocinar sopa de ratón

$$A := Cws$$

de la evidencia de que para hacer sopa de ratón hace falta el ingrediente que le dá nombre

$$B := \forall x(Cxs \rightarrow \exists y(Ry \wedge Iys)$$

y de la conciencia de que él es un ratón y no hay ningún otro

$$C := Rm \wedge \forall x(Rx \rightarrow x = m)$$

En el cálculo de primer orden es fácil deducirlo

$$\{A,\ B,\ C\} \vdash_{tab} D$$

A nadie se le escapa que la implicación personal condiciona la percepción de la conclusión y la acción futura tanto en el caso del ratón como en el del silogismo del reproche. Esos aspectos no se tienen en cuenta en la lógica clásica, pero tienen su hueco en otras lógicas epistémicas y autoepistémicas.

La lógica de primer orden contiene a la proposicional, pero es más potente. Hay que señalar las ventajas y los inconvenientes de la lógica clásica de primer orden, resaltando entre las primeras su expresividad y entre las segundas su mal comportamiento desde un punto de vista computacional: indecidibilidad para validez, difícil chequeo de modelos. Ello nos hace fijarnos en ciertos subconjuntos de las fórmulas de primer orden que son interesantes para nuestros propósitos. Destacan los conjuntos formados por aquellas fórmulas que emplean sólo relatores monarios, los que usan sólo un número finito de variables y los que tienen limitada la alternancia de cuantificadores; concretamente LPO^2 (lógica de primer orden con dos variables) es un fragmento decidible de LPO, aunque LPO^3 (con tres) ya es indecidible.

En el marco de la lógica clásica son especialmente sencillos los razonamientos lógicos que se resuelven con diagramas de Venn, su lenguaje tiene una complejidad que se sitúa entre la proposicional y la de primer orden de predicados monarios. En este contexto se presentan los silogismos.

3.4. Silogística

Aristóteles fue el primero que de manera sistemática trató con una cierta profundidad la relación que se establece entre las sentencias que forman parte de un razonamiento, observando que para estudiar la naturaleza de la deducción hace falta analizar primero la estructura de las que constituyen sus hipótesis y

su conclusión. En la lógica tradicional, de Aristóteles a Leibniz, incluso en Boole, ésta se toma de la gramática de las lenguas naturales[6]; es decir, una sentencia se analiza en términos de *sujeto* S y *predicado* P. Se distinguen cuatro *formas* típicas de proposiciones: A, E, I y O

El silogismo categórico es una estructura de proposiciones que se caracteriza:

1. Por tener dos premisas (mayor y menor) y una conclusión
2. Por tener sólo tres términos: Mayor P, medio M y menor S

Se llaman figuras del silogismo a las distintas posiciones que ocupa el término medio —el que desaparece de la conclusión— y que se expresan así:

Primera figura	Segunda figura	Tercera figura	Cuarta figura
$M \quad P$	$P \quad M$	$M \quad P$	$P \quad M$
$S \quad M$	$S \quad M$	$M \quad S$	$M \quad S$
$S \quad P$	$S \quad P$	$S \quad P$	$S \quad P$

El modo de un silogismo resulta de la combinación de las formas y figuras que contiene. Para cada figura hay sesenta y cuatro modos posibles. Como hay cuatro figuras, el resultado final es de 256 silogismos posibles. Por supuesto, no todos son válidos. La lógica tradicional seleccionaba de entre ellos a 24, a los que consideraba silogismos válidos, a muchos de los cuales se les atribuyeron nombres nemotécnicos en el medioevo. El más conocido es el de *BARBARA*, del que hacemos uso continuamente. Se trata de un silogismo de la primera figura en donde todos sus proposiciones son de la forma universal afirmativa, A. El ejemplo 2 enlaza varios argumentos simples, alguno de los cuales sigue aparentemente ese patrón

Me dijeron que no soy nadie. Nadie es perfecto.
Luego, yo soy perfecto.

Hoy no enseñamos la lógica aristotélica porque la lógica matemática la supera con creces; sin embargo debemos hacer justicia al filósofo y reconocer que la silogística es capaz de identificar adecuadamente los razonamientos correctos, algo que dista mucho de ser el caso en otras de sus aportaciones; por ejemplo, la física aristotélica. Por supuesto, no identifa a *todos* los correctos; estamos hablando de su reducido campo de aplicación, el de la lógica de predicados monarios y con una sola variable, una lógica que se sitúa entre la proposicional y la de primer orden. Para ser exactos, tampoco todos los seleccionados son razonamientos válidos con los estándares actuales. La razón es que para nosotros existe la cuantificación vacía y al afirmar que

Todos los misóginos son impresentables

no decimos que los haya, sino que caso de haberlos serían impresentables. En la interpretación de la silogíatica clásica se excluye la muy deseable situación de ausencia de misóginos.

[6] Nosotros ahora utlilizamos un análisis más rico, basado en la concepción de Frege.

3.5. Lenguajes clásicos

En la lógica clásica hay varias categorías de lenguajes: proposicional, de primer orden, de segundo orden, etc. Hemos visto que el de primer orden añade al proposicional la capacidad de analizar las fórmulas atómicas mediante relatores, functores y constantes y la cuantificación sobre individuos. Así que para la pregunta

¿Qué lenguaje necesitamos?

no hay una respuesta categórica, depende de para qué. Seguimos preguntándonos,

¿Se pueden expresar en primer orden todas las propiedades imaginables de las estructuras matemáticas?

El lenguaje de segundo orden añade al anterior la facultad de cuantificar sobre conjuntos y relaciones. Veamos algunos ejemplos

1. El *Axioma de Inducción* puede formularse del modo siguiente, y retener todo su poder expresivo:

$$\forall X(Xc \wedge \forall x(Xx \rightarrow X\sigma x) \rightarrow \forall x\ Xx)$$

   Esta fórmula dice: *Toda propiedad que valga para el cero y para el siguiente de cualquier número que la tenga, es una propiedad de todos los números.* La aritmética de Peano de segundo orden AP^2 la forman este axioma, junto al de *inyectividad* de la función del siguiente y la exigencia de que el *cero* no sea siguiente de ningún número.
2. La *Identidad entre Individuos* puede introducirse por definición y no ser, como en la lógica de primer orden, un concepto lógico, primitivo; es decir, tomado directamente de la metateoría. La definición comúnmente aceptada es la de Leibniz, que en *SOL* presenta el siguiente aspecto:

$$\forall xy(x = y \leftrightarrow \forall X(Xx \leftrightarrow Xy))$$

   Esta fórmula dice: *"Dos individuos son iguales si, y sólo si, comparten todas sus propiedades".*
3. El concepto intuitivo de *la mayoría de los* R *son* S —i.e., la mayor parte de los elementos que tienen la propiedad R tienen también la propiedad S—, puede expresarse en lógica de segundo orden con dos relatores monarios para R y S del modo siguiente:

$$\neg\exists X^2(\forall x(\exists y\ X^2xy \leftrightarrow Rx \wedge Sx) \wedge \forall x(\exists y\ X^2yx \rightarrow Rx \wedge \neg Sx)$$
$$\wedge\forall xyz(X^2xy \wedge X^2xz \rightarrow y = z) \wedge \forall xyz(X^2xy \wedge X^2zy \rightarrow x = z))$$

   Esta fórmula dice: *"no hay ninguna función inyectiva de* $R \cap S$ *en* $R - S$*".* Se acepta que esta formulación logra captar la idea intuitiva de: *"la mayor parte de los* R *son* S", puesto que está diciendo que el conjunto $R \cap S$ es *"mayor"* que el conjunto $R - S$.

Vemos que si nos preguntamos

¿Sirve la lógica de primer orden para axiomatizar toda la matemática?

la respuesta es que no. El lenguaje de la lógica de segundo orden es más expresivo que el de primer orden y éste que el de orden cero. Sin embargo, las propiedades lógicas de estos lenguajes van decreciendo: mientras que la lógica proposicional posee un cálculo deductivo correcto, completo y es decidible, la de primer orden posee un cálculo correcto y completo, pero ya no es decidible, y la de segundo orden ni es decidible ni posee un cálculo completo.

Conclusión 1 *Una lógica es como una balanza: en un platillo se pone el poder expresivo de la lógica y en el otro las propiedades lógicas. En la lógica proposicional pesan más las propiedades lógicas, en la de segundo orden la capacidad expresiva, mientras que la de primer orden está más equilibrada. Sabiendo esto somos nosotros los que decidiremos qué lógica necesitamos, qué virtudes nos interesa conservar.*

4. Argumentación y Retórica

La lógica también está emparentada con la Teoría de la Argumentación y con la Retórica. Los estudiantes de filosofía recordamos a los sofistas y también los diálogos platónicos en los que Sócrates termina acorralando a sus adversarios, con la fuerza de sus argumentos. La lógica proporciona los patrones o estrategias que se pueden desarrollar en el curso de una buena argumentación competitiva. Puesto que la mejor y más honesta manera de vencer en una controversia es proporcionar una argumentación correcta, el interés práctico conduce al teórico, investigándose la inferencia válida.

La teoría de la argumentación se centra en el análisis de la estructura de los argumentos y distingue la corrección formal de un argumento de su solidez, que tiene en cuenta el grado de justificación de las premisas. Por lo que respecta a los argumentos complejos se distingue entre argumentación *concatenada*, *coorientada* y *antiorientada*. La primera se caracteriza porque la conclusión de un argumento forma parte del conjunto de hipótesis del que lo sigue. De esta clase es la llevada a cabo en el ejemplo 2.

El discurso más importante en la antigüedad clásica era el discurso legal y el político y ahí lo principal era convencer, por lo que recoge también consideraciones psicológicas y lingüísticas. A nadie se les escapa que convencer y tener objetivamente la razón no son siempre la misma cosa. El filósofo Arthur Schopenhauer en su irónico opúsculo *El arte de tener razón* propone 38 estratagemas, todas *suculentas*, para vencer. Cito la última, que como veis se aplica continuamente en las tertulias radiofónicas y televisivas.

> "Cuando se advierte que el adversario es superior y que uno nunca conseguirá llevar razón, persolalícese, séase ofensivo, grosero".

El objetivo no es otro que vencer, de la manera que sea

> "La dialéctica **erística** es el arte de discutir, y de discutir de tal modo que uno siempre lleve razón, es decir, **per fas et nefas** [justa o injustamente]"

Las estrategias de Schopenhauer tendrían una traducción en nuestro mundo digital y en vez de ser ofensivo o grosero se podría destruir archivos, hacer que la aplicación del usuario entre en un bucle sin fin, modificar su fondo de escritorio sin previo aviso, etc.

5. Otras lógicas

En el ejemplo 7 lo primero que deberíamos hacer es identificar premisas y conclusión; aquí vemos que la pregunta inicial es justamente la conclusión del argumento. Conociendo los recursos de la retórica, la razón que parece sustentarlo es la de la permanencia de las leyes que gobiernan el cosmos, en especial respecto a la existencia de objetos; el argumento contenido en este pasaje, podría reformularse así:

Hipótesis 1 $\alpha :=$ *Si existen cosas en un momento dado, entonces en todo momento anterior han existido cosas*
Hipótesis 2 $\beta :=$ *Existen cosas hoy*
Hipótesis 3 $\gamma :=$ *El principio de los tiempos es anterior a todo*

LUEGO
Conclusión $\delta :=$ *Existían cosas al inicio de los tiempos*

Utilizando el siguiente lenguaje formal de primer orden:

$Exy :=$ *y existe en el momento x*
$Cx :=$ *x es una cosa*
$Mx :=$ *x es un momento de tiempo*
$Axy :=$ *x es anterior a y*
$a :=$ *principio de los tiempos*
$h :=$ *hoy*

Escribiríamos:

$$\alpha := \forall y(My \wedge \exists x(Cx \wedge Exy) \rightarrow \forall z(Mz \wedge Azy \rightarrow \exists x(Cx \wedge Exz))) \qquad (1)$$
$$\beta := \exists x(Cx \wedge Exh) \qquad \gamma := \forall y(My \rightarrow Aay) \qquad \delta := \exists x(Cx \wedge Exa)$$

La interpretación de estas fórmulas contará con un universo heterogéneo $\mathcal{U}$ formado tanto por cosas como por momentos de tiempo y varios relatores. En este caso lo natural sería tener dos universos $\mathcal{U}_1$ e $\mathcal{U}_2$ —para cosas uno y para momentos otro— y prescindir de los relatores monarios, al tener distintos tipos de variables para cuantificar sobre cada universo. Esto es lo que se hace en lógica multivariada o heterogénea.

5.1. Lógica multivariada

En muchas de las ramas de la matemática, de la filosofía, de la I.A. y de la informática formalizamos enunciados relativos a diversos tipos de objetos. Por consiguiente, tanto los lenguajes lógicos utilizados, como las estructuras matemáticas que los interpretan son multivariadas o heterogéneas; esto es, el conjunto de las variables del lenguaje toma valores sobre diversos universos o dominios. Son numerosos los ejemplos de materias que utilizan fórmulas y estructuras multivariadas: En *geometría*, por tomar un ejemplo clásico y sencillo, usamos distintos universos para *puntos*, *líneas*, *ángulos*, *triángulos*, etc. En *computación* utilizamos invariablemente estructuras multivariadas: lo típico es tener universos de *datos*, *números naturales* y *operadores booleanos*. Podemos añadir otros para *números reales*, *cadenas de caracteres*, *matrices*, etc.

¿Qué lenguaje y qué lógica es el adecuado en cada uno de estos campos?

La respuesta es que la lógica multivariada es la que mejor les cuadra.

Éste es también el caso en nuestro ejemplo 7. Podemos simplificar la formalización al emplear un *lenguaje bivariado* con dos clases de variables para *cosas* y para *instantes de tiempo*, un predicado binario de *existencia en un instante* dado y dos constantes temporales, *hoy* y el *inicio de los tiempos*. En él escribiríamos:

$$\alpha := \forall t(\exists x Ext \rightarrow \forall t\prime(At\prime t \rightarrow \exists x Ext\prime) \qquad \beta := \exists x Exh \qquad (2)$$
$$\gamma := \forall t\ Aat \qquad \delta := \exists x Exa$$

Comparando las formalizaciones en lógica de primer orden 1 y en multivariada 2 se entiende bien la relación entre ambas lógicas. La reducción de la lógica multivariada a la univariada es un resultado no sólo bien conocido desde antiguo, sino también el planteamiento que normalmente se hace en los libros de texto. El proceso se lleva a cabo a dos niveles: hay una *traducción sintáctica* de las fórmulas multivariadas a las univariadas —conocida como *relativización de cuantificadores*— y una *conversión semántica* de estructuras —conocida como *unificación de dominios*—.

Para traducir tomamos un lenguaje de primer orden sin variedades —esto es, con una sola clase de variables— con los mismos signos de operación que tuviéramos en la multivariada y le añadimos tantos relatores monarios como variedades hubiera. Cada fórmula cuantificada sobre una variedad i

$$\forall x^i \varphi(x^i)$$

será reemplazada por una fórmula cuantificada condicional, en cuyo antecedente decimos sobre qué variedad se restringe la cuantificación

$$\forall x(Q^i x \rightarrow \varphi(x)^*)$$

La nueva estructura univariada obtenida mediante unificación de dominios tendrá un solo universo constituido por la unión de todos los universos de la que se reduce,

las relaciones de la estructura multivariada pasan a serlo de la nueva univariada y las funciones de la multivariada se extienden para que puedan serlo de la univariada, añadiendo valores arbitrarios para los nuevos elementos.

Volvamos al caso concreto de nuestro ejemplo 7. Aunque hemos simplificado la formalización, no termina de ser completamente natural. Aquí lo natural sería emplear la lógica temporal o, aún mejor, lógica híbrida.

5.2. Lógica temporal

En ella añadimos al lenguaje proposicional o de primer orden nuevos operadores modales:

$\langle P\rangle\,\varphi$	Alguna vez en el pasado, φ
$\langle F\rangle\,\varphi$	Alguna vez en el futuro, φ
$[P]\,\varphi$	Siempre en el pasado, φ
$[F]\,\varphi$	Siempre en el futuro, φ

Prior (1967) formula varias tesis sobre el tiempo que servirán de axiomas en su lógica temporal; entre ellas cabe destacar:

- $[F]\,\varphi \rightarrow \langle F\rangle\,\varphi$ *Lo que siempre será verdadero, lo será alguna vez*
- $\langle F\rangle\,\varphi \rightarrow \langle F\rangle\,\langle F\rangle\,\varphi$ *Si φ será verdadero alguna vez, entonces será alguna vez verdadero que φ será verdadero alguna vez*

El pasado y el futuro son interdefinibles

- $\varphi \rightarrow [P]\,\langle F\rangle\,\varphi$ *Lo que es verdadero, fue siempre verdadero que alguna vez sería verdadero*
- $\varphi \rightarrow [F]\,\langle P\rangle\,\varphi$ *Lo que es verdadero, será siempre verdadero que alguna vez fue verdadero*

La lógica híbrida se da cuenta de las incongruencias de la lógica modal, en donde, por una parte los estados o momentos de tiempo son cruciales pero no podemos referirnos a ellos ya que el lenguaje carece de los elementos necesario. En lenguajes híbridos se puede introducir referencias explícitas a los elementos del dominio de un modelo; por ejemplo, momentos específicos (días años, etc.). De esta forma podemos mejorar el poder expresivo, modelar indexicales temporales (ayer, hoy, mañana, ahora,...) y definir propiedades relevantes para la lógica temporal (irreflexividad, asimetría). La lógica híbrida también posee una elegante teoría de la prueba, próxima a los sistemas deductivos etiquetados de Gabbay.

¿Cómo se crea una lógica híbrida?

Prior en sus últimos trabajos utiliza términos para referirse a instantes de tiempo y los trata como fórmulas. Así

$$\langle F\rangle\,(i \wedge p)$$

dice que el instante i y la proposición p están en el futuro y coinciden. Esto es, p ocurre en el instante futuro i. Para ello hemos de extender el lenguaje añadiendo a los átomos un conjunto de nominales

$$\text{ATOM} \cup \text{NOM}$$

la idea es usar fórmulas para referirse a los instantes. También agregamos un conjunto de operadores modales

$$\{@_i \mid i \in \text{NOM}\}$$

que nos permite formar nuevas fórmulas $@_i\varphi \in \text{FORM}$ para indicar que en el instante i la fórmula φ es verdadera. Debemos interpretar $i \in \text{NOM}$ en $\mathcal{A}$ como una clase unitaria: $i^{\mathcal{A}} = \{a\}$ y definir la interpretación de las fórmulas con nuevos operadores

$$\mathcal{A}, w \Vdash @_i\varphi \qquad \text{syss} \qquad \mathcal{A}, i^{\mathcal{A}} \Vdash \varphi$$

En la lógica híbrida escribimos nuestro argumento del jade celeste de forma más simple:

$$\begin{array}{ll} \text{Hipótesis} & \alpha := q \to [P]\, q \qquad \beta := @_h\ q \qquad \gamma := a \to [P]\, \bot \\ \text{Conclusión} & \delta := @_a\ q \end{array}$$

Esta conclusión es deducible en lógica híbrida, pues ella cuenta entre sus axiomas el que establece que la relación temporal conecta fuertemente a dos instantes cualesquiera

$$@_a\ h \vee @_a\ \langle P\rangle\, h \vee @_h\ \langle P\rangle\, a$$

La prueba procedería así:

- Primer caso: $@_a\ h$. Usando la premisa $@_h\ q$ concluimos que $@_a\ q$
- Segundo caso: $@_a\ \langle P\rangle\, h$. Usando la premisa $a \to [P]\, \bot$ obtenemos $@_a\ [P]\, \bot$ (La regla del operador $@_a$ nos permite escribir $@_a(a \to [P]\, \bot)$, la de reflexividad de la identidad nos da $@_a\ a$ y ahora basta emplear el axioma K y aplicar la regla de Modus Ponens). Mediante la regla de Binding aplicada a $@_a\ \langle P\rangle\, h$ y $@_a\ [P]\, \bot$ obtenemos lo deseado.
- Tercer caso: $@_h\ \langle P\rangle\, a$. Usando las premisas $q \to [P]\, q$ y $@_h\ q$ obtenemos $@_h\ [P]\, q$ por el mismo procedimiento que en el caso anterior. Ahora Binding aplicada a $@_h\ \langle P\rangle\, a$ y $@_h\ [P]\, q$ produce la conclusión deseada $@_a\ q$.

Otros ejemplos nuestros también se formulan bien en lógica híbrida. El argumento del primer niño del pasaje de Confucio 8 podría reformularse así:

Hipótesis 1 $\alpha :=$ *Cuando el sol sale es grande*
Hipótesis 2 $\beta :=$ *A mediodía el sol es pequeño*
Hipótesis 3 $\gamma :=$ *Lo grande está cerca y lo pequeño lejos*

LUEGO
Conclusión $\delta :=$ *Cuando el sol sale está cerca y a mediodía, lejos*

Necesitamos un lenguaje híbrido de primer orden en donde poder expresar las propiedades de ser grande y pequeño y estar cerca o lejos. Un nombre para el sol y dos nominales para el alba y el mediodía. Podemos emplear éste

$$\begin{array}{l} Gx := x \text{ es grande} \\ Px := x \text{ es pequeño} \\ Cx := x \text{ está cerca} \\ Lx := x \text{ está lejos} \end{array}$$

$$\begin{array}{l} m := \text{mediodía} \\ s := \text{sol} \\ a := \text{alba} \end{array}$$

Como veis, la formalizadción en lógica híbrida es simple y expresiva

$$\alpha := @_a\ Gs \qquad \beta := @_m Ps \qquad \gamma := \forall x(Gx \to Cx) \wedge \forall x(Px \to Lx)$$
$$\delta := @_a\ Cs \wedge @_m\ Ls$$

6. Procedimientos de búsqueda

Aquí no se trata de comprobar si un enunciado es consecuencia de un conjunto de hipótesis, sino de encontrar una solución *"razonable"* a un problema o una explicación *"convincente"* de un hecho sorprendente.

6.1. Lógica abductiva

En el ejemplo 4 la conclusión A es que el Dr Watson ha estado en Afganistán. Sherlock se basa en la evidencia de que está herido y moreno $H \wedge M$. Su razonamiento basado en su experiencia y conocimientos le indica que el Dr ha viajado al trópico, ya que está moreno y éste no es su color natural (tampoco fácil de conseguir con el clima británico). Por otra parte, sabe que es médico militar y que está herido y que en Afganistán su país mantiene una contienda pues tanto el imperio ruso como el británico se lo disputan. Aparentemente ha concluido A de la premisa $H \wedge M$ usando una hipótesis $A \to H \wedge M$ que parece plausible y compatible con el resto de sus conocimientos.

$$\frac{\begin{array}{c} H \wedge M \\ A \to H \wedge M \end{array}}{A}$$

Esto en lógica clásica es una aberración, la conocida como *falacia de la afirmación del consecuente*. Sin embargo, en lógica abductiva el planteamiento es distinto. A pretende ser una explicación de un hecho sorprendente $H \wedge M$ compatible con sus conocimientos sobre el caso. Por supuesto, no hay certeza absoluta para A, y en esto se distingue claramente del caso deductivo.

La interpretación más usual de la abducción como inferencia lógica es como *deducción para atrás más condiciones adicionales*, según Aliseda. El razonamiento abductivo se realiza en base a una cierta teoría Γ y a una fórmula φ que ha

de ser explicada, para la que se postula una explicación α. Debe suceder que de la explicación sea consecuencia lógica el hecho φ, módulo la teoría

$$\Gamma \cup \alpha \models \varphi$$

pero además la explicación debe ser consistente con la teoría $\Gamma \cup \alpha \not\models \bot$ y ser mínima, lo que significa que su forma lógica es muy simple. Una condición adicional, para que el hecho sorprenda, es que de la teoría ni él ni su negación se deduzcan.

$$\Gamma \not\models \varphi \qquad \Gamma \not\models \neg\varphi$$

Un hecho sorprende por su novedad o su anomalía, en el primer caso ni ese hecho ni su negación se desprende de la teoría, en el segundo se trata de que en verdad la teoría apoya su contrario $\Gamma \models \neg\varphi$. Las operaciones abductivas de cambio epistémico están asociadas a las propiedades mencionadas. En particular,

- la *expansión abductiva* que dada una novedad abductiva φ —tal que $\Gamma \not\models \varphi$ y $\Gamma \not\models \neg\varphi$— y una explicación consistente α —tal que $\Gamma \cup \alpha \models \varphi$— nos permite añadir a la teoría tanto la novedad como la explicación $\Gamma \cup \varphi \cup \alpha$
- la *revisión abductiva* que dada una anomalía abductiva φ —tal que $\Gamma \not\models \varphi$ y $\Gamma \models \neg\varphi$— una explicación consistente α se calcula a partir de un subconjunto Δ de Γ obtenido mediante revisión. Básicamente hacemos que $\Delta \not\models \neg\varphi$ al quitarle algunas de sus fórmulas y luego obtenemos la explicación consistente de la forma usual, pero usando la teoría revisada.

6.2. Soluciones *"razonables"*

El cálculo de enunciados subayace al razonamiento cotidiano, lo empleamos para hallar soluciones a nuestros problemas, como el Padrino en el caso que sigue:

Example 9. Robo de Archivos de la MAFIA

Al llegar el Padrino a su despacho notó que alguien había entrado en él, ¡incluso habían revuelto sus archivos! Pudo comprobar que faltaban algunos documentos comprometedores.
La investigación del caso arroja estos datos:

$A :=$ Nadie más que P, Q y R están bajo sospecha y al menos uno es traidor.
$B :=$ P nunca trabaja sin llevar al menos un cómplice.
$C :=$ R es leal.

Contamos sólo con las hipótesis. Lo primero que hacemos es comprobar si el conjunto es *satisfacible* (consistencia semántica) ya que en caso contrario cualquier fórmula se seguiría de él y esas no serían soluciones razonables. Cuando el conjunto de hipótesis encierra contradicciones, para cualquier fórmula F : tanto F como $\neg F$ sería consecuencia del conjunto. En el caso del problema planteado, si el conjunto de hipótesis fuera insatisfacible, se podría demostrar para cada

uno de los implicados que es culpable. E incluso también la culpabilidad de cualquier otra persona, aunque no haya aparecido en el enunciado del problema. Más dramático aún, se demostraría que *tú* eres culpable. Bien es verdad que también se puede demostrar que todos somos inocentes como pajarillos...

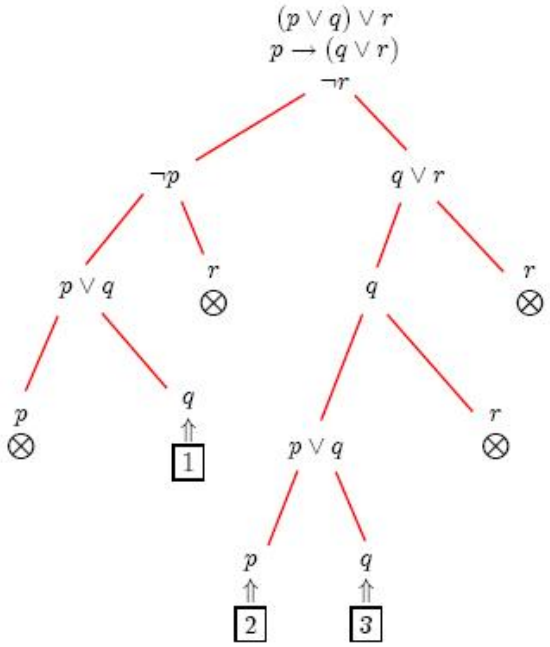
¡Descartamos por lo tanto las consecuencias de los conjuntos insatisfacibles!

Si el conjunto es satisfacible usaremos las ramas abiertas del tableau para hallar la solución: tomaremos los literales de dichas ramas y haremos la intersección; esto es, seleccionaremos los que están en todas las ramas, sólo esos. La conjunción de las fórmulas así obtenidas será la solución. Por supuesto, habrá otras muchas fórmulas que también serán consecuencia de las hipótesis, de hecho, infinitas. Hemos elegido ésta por ser la más simple. En los casos en los que la intersección de estos conjuntos de literales sea vacía, sigue habiendo fórmulas que son consecuencia de las hipótesis, nosotros hemos decidido descartarlas por no ser suficientemente contundentes.

Partimos del conjunto de hipótesis

$$\{(p \vee q) \vee r,\ p \rightarrow (q \vee r)\,,\ \neg r\}$$

y queremos obtener una solución *razonable*. Para comprobar que son compatibles hacemos un árbol.



$\{A, B, C\}$ es satisfacible pues hay 3 interpretaciones que hacen a A, B y C simultáneamente verdaderas, basadas en los conjuntos

Hay tres ramas abiertas de las que extraemos el conjunto de sus literales

$$\boxed{1} = \{\neg p, \neg r, q\} \qquad \boxed{2} = \{p, q, \neg r\} \qquad \boxed{3} = \{q, \neg r\}$$

Para hallar la conclusión hacemos su intersección

$$\boxed{1} \cap \boxed{2} \cap \boxed{3} = \{q, \neg r\}$$

La conclusión resultante es: $q \wedge \neg r$

7. Ejercicios para resolver

Exercise 1. **El Péndulo de Foucault**

Mediante este experimento, Foucault pretende demostrar la rotación de la Tierra.
Parte de una serie de observaciones experimentales, utiliza sus conocimientos de física y llega a la conclusión de que efectivamente la tierra rota.
El experimento consta de un péndulo colgado de un punto fijo (idealmente, el del **"plano de las estrellas fijas"**) al que se le hace oscilar libremente. Alrededor del péndulo, en el plano de la Tierra, se dibuja un círculo formado por pivotes (perfectamente calculado según las coordenadas geográficas del lugar). Foucault observó que en su oscilación, el péndulo iba derribando sucesivamente todos los pivotes. El sabía que el plano de oscilación del péndulo se mantenía fijo respecto del **"plano de las estrellas fijas"**. (La física establece que el plano de oscilación de cualquier péndulo se mantiene fijo respecto al plano sobre el que oscila.)

1. Formalizad en lógica proposicional el argumento que posiblemente utilizó Foucault usando las claves siguientes:

   $p :=$ gira el plano del péndulo respecto del plano de la Tierra
   $q :=$ gira el plano del péndulo respecto del plano de las estrellas fijas
   $t :=$ gira el plano de la Tierra respecto del de las estrellas fijas
   $s :=$ caen más de dos pivotes

   Reconstruyamos el razonamiento de Foucault. Vamos a emplear cuatro hipótesis. Os voy a dar unas pistas y vostros escribireis las fórmulas empleando el lenguaje proposicional. Entre las hipótesis hay hechos y leyes físicas
   $A :=$? (observa los pilotes que derriba)
   :
   $B :=$? (el plano de oscilación del péndulo se mantiene fijo)
   $C :=$? (relación existente entre el giro del plano del péndulo respecto de la tierra y la caída de pivotes; se descartan otras posibilidades, el experimento está "vigilado")

$D :=$? (expresa la relación entre los tres planos)
Conclusión:
$E := t$

2. Queremos demostrar que no existe ninguna interpretación $\Im$ en la que todas las hipótesis sean verdaderas sin que lo sea t también. O, lo que es equivalente, que $\{A,\ B,\ C,\ D\} \vdash E$

Exercise 2. Volvamos al ejemplo de la sopa de ratón. En el cálculo de primer orden es fácil deducir la conclusión. Demostrad que

$$\{A,\ B,\ C\} \vdash_{tab} D$$

Donde

$$A := Cws \qquad B := \forall x(Cxs \rightarrow \exists y(Ry \wedge Iys) \qquad C := Rm \wedge \forall x(Rx \rightarrow x = m)$$
$$D := Ims$$

Exercise 3. Volvamos al ejemplo de Confucio 8. Vamos a emplear el lenguaje clásico de primer orden. Necesitamos un lenguaje en el que poder expresar las propiedades de ser grande y pequeño en un momento dado y también relativizar las distancias. Podemos emplear éste

$Gxy :=$ x es grande en el momento y
$Pxy :=$ x es pequeño en el momento y
$Cxy :=$ x está cerca en el momento y
$Lxy :=$ x está lejos en el momento y
$Ox :=$ x es un objeto
$Mx :=$ x es un momento de tiempo

$m :=$ mediodía
$s :=$ sol
$a :=$ alba

1. La formalización de las hipótesis y de la conclusión en este lenguaje formal de primer orden resulta un poco farragosa, especialmente la tercera hipótesis. Hacedlo:

$\alpha :=$?
$\beta :=$?
$\gamma :=$?
$\delta :=$?

2. Para demostrar en el cálculo de tableaux que efectivamente $\{\alpha, \beta, \gamma\} \vdash \delta$ añadimos a las anteriores hipótesis las de *sentido común del uso de la lengua*

$$\zeta := Os \wedge Ma \wedge Mm$$

Resolvedlo en el cálculo de tableaux de primer orden; esto es, demostrad

$$\left\{\begin{array}{l}\alpha\\ \beta\\ \gamma\\ Os \wedge Ma \wedge Mm\end{array}\right\} \vdash_{tab} \delta$$

Referencias

1. Atocha Aliseda [2005] *Abductive Reasoning: Logical Investigationes into Discovery and Explanation.* Springer/Kluwer.
2. Carlos Areces.[2000] *Logic Engineering. The Case of Description and Hybrid Logics.* Ph.D. Thesis, Institute for Logic, Language and Computation, University of Amsterdam, . ILLC Dissertation Series 2000–5.
3. Patrick Blackburn, Maarten de Rijke, and Yde Venema. [2001]. *Modal Logic*, Cambridge University Press.
4. Dov Gabbay y Guenthner, F. editores. [2001]. *Handbook of Philosophical Logic* 2^{nd} edition. Kluwer Academic Publishers. Dordrecht. Holanda. vol 1 a 4. Segunda edición, 18 volúmenes en preparación.
5. Dov Gabbay. [1996]. *Labelled Deductive Systems.* Oxford University Press.
6. Manzano, M. [1996]. *Extensions of first order logic***.** Cambridge University Press.
7. Manzano, M. ed. [2004]. *Summa logicae en el siglo XXI.* Ediciones Universidad de Salamanca. (también en http://logicae.usal.es)
8. María Manzano y Antonia Huertas [2004]. *Lógica para principiantes.* Alianza Editorial
9. Johan van Benthem [2006]. *"Adiós a la soledad: modas dinámicas en la lógica actual"*AZAFEA.

Intelligent Behavior: Lessons from AI Planning

Héctor Geffner

ICREA & Universitat Pompeu Fabra.
Paseo de Circunvalación, 8. 08003 Barcelona, Spain
hector.geffner@upf.edu

Abstract. Humans encounter a huge variety of problems which they must solve using general methods. Even simple problems, however, become computationally hard for general solvers if the structure of the problems is not recognized and exploited. Work in Artificial Intelligence Planning and Problem Solving has encountered a similar difficulty, leading in recent years to the development of well-founded and empirically tested techniques for recognizing and exploiting structure, focusing the search for solutions in certain cases, and bypassing the need to search in others. These techniques include the automatic derivation of heuristic functions, the use of limited but effective forms of inference, and the compilation of domains, all of which enable a general problem solver to 'adapt' automatically to the task at hand. In this paper, I present the ideas underlying these new techniques, and argue for their relevance to models of natural intelligent behavior. The paper is not a review of AI Planning – a diverse field with a long history – but a personal appraisal of some recent developments and their potential bearing on accounts of action selection in humans and animals.

1 Introduction

In the late 50's, Newell and Simon introduced the first AI planner – the General Problem Solver or GPS – as a psychological theory [1, 2]. Since then, Planning has remained a central area in AI while changing in significant ways: it has become more mathematical (a variety of planning problems has been clearly defined and studied) and more empirical (planners and benchmarks can be downloaded freely, and competitions are held every two years), and as a result, new ideas and techniques have been developed that enable the automatic solution of large and complex problems [3].

AI Planning studies languages, models, and algorithms for describing and solving problems that involve the selection of actions for achieving goals. In the simplest case, in *classical planning*, the actions are assumed deterministic, while in *contingent planning,* actions are non-deterministic and there is feedback. In all cases, the task of the planner is to compute a plan or solution; the *form* and *cost* of these solutions depending on the model; e.g., in classical planning, solutions are sequences of actions and cost is measured by the number of actions, while in planning with uncertainty and feedback, solutions map states into actions, and cost stands for expected or worst-possible cost.

Planning is a form of 'general problem solving' over a class of models, or more precisely, a *model-based* approach to intelligent behavior: given a problem in the form of a

A. Fernández-Caballero & S. Miguel (Eds.): 50 Años de la Inteligencia Artificial, pp. 91-102, 2006.

compact description of the actions, sensors (if any), and goals, a planner must compute a solution, and if required, a solution that minimizes costs. Some of the models used in planning, as for example Markov Decision Processes (MDPs), are not exclusive to AI Planning, and are used for example in Control Theory [4], Reinforcement Learning [5], and Behavioral Ecology [6, 7] among other fields. What is particular about AI planning are the *languages* for representing these models, the *techniques* for solving them, and the ways these techniques are *validated.* Techniques do matter quite a lot: even simple problems give rise to very large state spaces that cannot be solved by exhaustive methods. Consider the well known Rubik Cube puzzle: the number of possible configurations is in the order of trillions, yet methods are known for solving it, even optimally, from arbitrary configurations [8]. The key idea lies in the use of *admissible heuristic functions* that provide an optimistic approximation of the number of moves to solve the problem from arbitrary configurations. These functions enable the solution of large problems, even ensuring optimality, by focusing the search and avoiding most states in the problem. Interestingly, recent work in planning has shown that such functions can be derived *automatically* from the problem description [9], and can be used to drive the search in problems involving uncertainty and feedback as well [10]. Such functions can be understood as a specific and robust form of *means-ends analysis* [1, 2] that produces goal-directed behavior in complex settings even in the presence of large state and action spaces.

In this paper, we review some of the key computational ideas that have emerged from recent work in planning and problem solving in AI, and argue that these ideas, although not necessarely in their current form, are likely to be relevant for understanding natural intelligent behavior as well. Humans encounter indeed a huge variety of problems which they must solve using general methods. It cannot be otherwise, because there cannot be as many methods as problems. Yet, simple problems become computationally hard for a general solver if the structure of the problems is not recognized and exploited. This is well known in AI, where systems that do not exhibit this ability tend to be shallow and brittle. In the last few years, however, work in Planning and Problem Solving has led to well-founded and empirically tested techniques for recognizing and exploiting structure, focusing the search for solutions, and in certain cases, bypassing the need to search altogether. These techniques include the automatic derivation of heuristic functions, the use of limited but effective forms of inference, and the compilation of domains, all of which enable a general problem solver to 'adapt' automatically to the task at hand. Interestingly, the need for focusing the search for solutions has been recognized in a number of recent works concerned with natural intelligent behavior, where it has been related to the role of emotions in the appraisal and solution of problems. We will say more about this as well.

Since Newell's and Simon's GPS, the area of AI planning has departed from the original motivation of understanding human cognition to become the mathematical and computational study of the problem of selecting actions for achieving goals. After all these years, however, and given the progress achieved, it is time to reflect on what has been learned in the abstract setting, and use it for informing our theories in the natural setting. This exercise is possible and may be quite rewarding. It parallels the approach advocated by David Marr, and echoed more recently by [11] and others in

the Brain Sciences; namely: characterize *what* needs to be computed, *how* it can be computed, and how these computations are *approximated* in real-brains. The findings that we summarize below, aim to provide a partial account of the first two tasks.

A few methodological comments before proceeding. First about *domain-generality* vs. *domain-specificity* in action selection and problem solving. I have said that humans are capable of solving a wide range of problems using general methods. This, however, is controversial. Both evolutionary psychologists [12] and cognitive scientists from the 'fast and frugal heuristics' school [13] place an emphasis on modularity and domain-specificity. Others, without necessarily denying the role of specialization, postulate the presence of general reasoning and problem solving mechanisms as well, at least in humans (see for example [14]). We are not going to address this controversy here, just emphasize that 'general' and 'adapted' are not necessarily opposite of each other. Indeed, the work in AI planning is domain-independent, yet the recent techniques illustrate how a general problem solver can 'adapt' to a specific problem by recognizing and exploiting structure, for example, in the form of heuristic functions. These heuristics are indeed in line with the 'fast and frugal heuristics', the difference being that they are general and can be extracted automatically from problem descriptions.

Another distinction that is relevant for placing the work in AI Planning within the broader work on Intelligent Behavior is the one between *finding solutions* vs. *executing solutions.* For many models, such as those involving uncertainty and feedback, the solutions, from a mathematical point of view, are functions mapping states into actions (these functions are called *closed-loop policies*, and in the partially observable case map actually *belief states* into actions; see below). These functions can be represented in many ways; e.g. as condition, action rules, as value functions, etc. Indeed, in what is often called *behavior-based AI* [15], these solutions are encoded by hand for controlling mobile robots. In nature, similar solutions are thought to be encoded in brains but not by hand but by evolution. Representing and executing solutions, however, while challenging, is different than coming up with the solutions in the first place which is what AI Planning is all about. Whether this is a requirement of intelligent behavior in animals is not clear although it seems to be a distinctive feature of intelligent behavior in humans. Interestingly, in many cases, the same models can be used for both *understanding* the solutions found in nature, and for *generating* those solutions [16]. The interest in the latter case, however, is not only with the models but also with the algorithms needed for solving those models effectively. We thus consider both models and algorithms.

2 Models

Most models considered in AI Planning can be understood in terms of *actions* that affect the *state* of a system, and can be given in terms of

1. a discrete and finite state space S,
2. an initial state $s_0 \in S$,
3. a non-empty set of terminal states $S_T \subseteq S$,
4. actions $A(s) \subseteq A$ applicable in each non-terminal state,
5. a function $F(a, s)$ mapping non-terminal states s and actions $a \in A(s)$ into *sets* of states

6. action costs $c(a, s)$ for non-terminal states s, and
7. terminal costs $c_T(s)$ for terminal states.

In deterministic planning, there is a single predictable next state and hence $|F(a, s)| = 1$, while in non-neterministic planning $|F(a, s)| \geq 1$. In addition, in probabilistic planning (MDPs), non-deterministic transitions are weighted with probabilities $P_a(s'|s)$ so that $\sum_{s' \in F(a,s)} P_a(s'|s) = 1$. In general, action costs $c(a, s)$ are assumed to be positive, and terminal costs $c_T(s)$ non-negative. When zero, terminal states are called *goals*. The models underlying 2-player games such as Chess can be understood also in these terms with opponent moves modeled as non-deterministic transitions. Often models are described in terms of rewards rather than costs, or in terms of both, yet care needs to be taken so that models have well-defined solutions. State models of this type are also considered in Control Theory [4], Reinforcement Learning [5], and Behavioral Ecology [6, 7]. In [17], it is shown how problems involving partial feedback can be reformulated as problems involving full state feedback over *belief states*; i.e., states that encode the information about the true state of the system. All these problems can also be cast as *search problems* in either the original state space or belief space [10].

The solutions to these various state models have a mathematical form that depends on the type of feedback. In problems without feedback, solutions are sequences of actions, while in problems with full-state feedback solutions are functions mapping states into actions (called also closed-loop control policies). The form of the solution to the various models need to be distinguished from the way they are represented. A common, compact representation of policies is in terms of condition, action rules; yet many of the standard algorithms assume a representation of policies in terms of less-compact value functions. The problem of combining robust algorithms with compact representations is not yet solved, although significant progress has been achieved when actions can be assumed to be deterministic.

From a complexity point of view, if there are n variables, the state space (range of possible value assignments) is exponential in n. Thus, except for problems involving very few variables, exhaustive approaches for specifying or solving these models are unfeasible. A key characteristic of AI Planning are the languages for representing these models, and the techniques used for solving them.

3 Languages

A standard language for representing state models in compact form is Strips [18].[1] In Strips, a problem P is expressed as a tuple $P = \langle A, O, I, G \rangle$ where A is the set of atoms or boolean variables of interest, O is the set of actions, and $I \subseteq A$, and $G \subseteq A$ are the atoms that are true in the initial and goal situations respectively. In addition, each action $a \in O$ is characterized by three sets of atoms: the atoms $pre(a)$ that must be true in order for the action to be executable (preconditions), the atoms $add(a)$ that become true after the action is done (add list), and finally, the atoms $del(a)$ that become false after doing the action (delete list).

[1] Strips is the name of a planner developed in the late 60's at SRI, a successor of Newell's and Simon's GPS.

A Strips planner is a *general problem solver* that accepts descriptions of arbitrary problems in Strips, and computes a solution for them; namely, sequences of actions mapping the initial situation into the goal. Actually, any deterministic state model can be expressed in Strips, and any Strips problem $P = \langle A, O, I, G \rangle$ defines a precise state model $S(P)$ where

- the states s are the different subsets of atoms in A
- the initial state s_0 is I
- the goal states s_G are those for which $G \subseteq s_G$
- $A(s)$ is the subset of actions $a \in O$ s.t. $pre(a) \subseteq s$
- $F(a, s) = \{s + Add(a) - Del(a)\}$, for $a \in A(s)$
- the actions costs $c(a, s)$ are uniform (e.g., 1)

Extensions of the Strips language for accommodating non-boolean variables and other features have been developed, and planners capable of solving large and complex problems currently exist. This is the result of new ideas and a solid empirical methodology in AI Planning following [19], [20], and others in the 90's.

4 Is Strips Planning relevant at all?

Before getting into the techniques that made this progress possible, let us address some common misconceptions about Strips planning. First, it is often said that Strips planning cannot deal with uncertainty. This is true in one way, but not in another. Namely, the model $S(P)$ implicit in a Strips encoding P does not *represent* uncertainty. Yet this does not imply that Strips planning cannot *deal* with uncertainty. It actually can. Indeed, the 'winner' of the ICAPS 2004 Probabistic Planning Competition [21], FF-Replan,[2] is based on a Strips planner called FF [22]. While the actions in the domain were probabilistic, FF-Replan ignores the probabilities and replans from scratch using FF after every step. Since currently, this can be done extremely fast even in domains with hundred of actions and variables, this deterministic re-planner did better than more sophisticated probabilistic planners. It does not take much to see that this strategy may work well in a 'noisy' Block Worlds domain where blocks may accidentally fall off gripper, and actually it is not trivial to come up with domains where this strategy will not work (this was indeed the problem in the competition). Control engineers know this very well: stochastic systems are often controlled by closed-loop control policies designed under deterministic approximations, as in many cases errors in the model can be safely corrected through the feedback loop.

A second misconception about Strips or 'classical' planning is that actions denote 'primitive operations' that all take a unit of time. This is not so: Strips planning is about planning with operators that can be chacterized in terms of pre and postconditions. The operator themselves can be abstractions of lower level policies, dealing with low level actions and sensors. For example, the action of grabbing a cup involves moving the arm in certain ways, sensing it, and so on; yet for higher levels, it is natural to assume that the action can be summarized in terms of preconditions involving the proximity of the cup, a free-hand, etc; and postconditions involving the cup in the hand and so

[2] FF-Replan was developed by SungWook Yoon, Alan Fern and Robert Givan from Purdue.

on. Reinforcement learning has been shown to be a powerful approach for learning low-level skills, but it has been less successful for integrating these skills for achieving high-level goals. The computational success of Strips planning suggests that one way of doing this is by characterizing low-level behaviors in terms of pre and postconditions, and feeding such behaviors into a planner.

5 Heuristic Search

How can current Strips planners assemble dynamically and effectively low-level behaviors, expressed in terms of pre and post conditions, for achieving goals? The idea is simple: they exploit the structure of the problems by extracting automatically informative heuristic functions. While the idea of using heuristic functions for guiding the search is old [23], the idea of extracting these functions automatically from problem encodings is more recent [24, 25], and underlies most current planners.

In order to illustrate the power of heuristic functions for guiding the search, consider the problem of looking in a map for the shortest route between Los Angeles and New York. One of the best known algorithms for finding shortest routes is Dijkstra's algorithm [26]: the algorithm efficiently and recursively computes the shortest distances $g(s)$ between the origin and the closest 'unvisited' cities s til the target is reached. A characteristic of the algorithm when applied to our problem, is that it would first find a shortest path from LA to Mexico City, even if Mexico City is way out of the best path from LA to NY. Of course, this is not the way people find routes in a map. The *heuristic search algorithms* developed in AI approach this problem in a different way, taking into account an estimate $h(s)$ of the cost (distance) to go from s to the goal. In route finding, this estimate is given by the Euclidian distance in the map that separates s from the goal. Using then the sum of the cost $g(s)$ to get to s and an estimate $h(s)$ of the cost-to-go from s to the goal, heuristic or informed search algorithms are much more *focused* than blind search algorithms like Dijkstra, without sacrificing optimality. For example, in finding a route from Los Angeles to New York, heuristic search algorithms like A* or IDA* [27, 28], would never consider 'cities' whose value $g(s) + h(s)$ is above the real cost of the problem. These algorithms guarantee also that the solutions found are optimal provided that the heuristic function h is *admissible* or *optimistic*, i.e., if for any s, $h(s) \leq V^*(s)$, where V^* is the optimal cost function. In the most informed case, when $h = V^*$, heuristic search algorithms are completely focused and consider only states along optimal paths, while in the other extreme, if $h = 0$, they consider as many states as Dijkstra's algorithm. Most often, we are not in either extreme, yet good informed heuristics can be found that reduce the space to search quite drastically. For example, while with today's technology it is possible to explore in the order of 10^{10} states, optimal solutions to arbitrary configuration of the Rubik's Cube with more than 10^{20} states, have been reported [8]. These search methods are very selective and consider a tiny fraction of the state space only, smaller actually than $1/10^{10}$.

6 Deriving Heuristic Functions

Two key questions arise: 1) How can these heuristics be obtained? and 2) Whether similar gains can be obtained in other models, e.g., when actions are not deterministic and states are not necessarily fully observable. We address each question in turn.

The power of current planners arises from methods for extracting heuristic values $h(s)$ automatically from problem encodings. The idea is to set the estimated costs $h(s)$ of reaching the goal from s to the cost of solving a simpler, relaxed problem. Strips problems, for example, can be relaxed by dropping the delete lists. Solving (non-optimally) a delete-free Strips problem can be done quite efficiently, and the heuristic $h(s)$ can be set to the cost of the relaxation. The idea of obtaining heuristics by solving relaxed problems is old [27], but the use of Strips relaxations for deriving them automatically for planning is more recent [24, 25]. Since then other relaxations have been been considered. In [25], the derived heuristics are used for selecting actions greedily, in real-time, without finding a complete plan first. The proposal is closely related to the *spreading activation model* of action selection in [29], with *activation levels* replaced by or interpreted as *heuristic values* (cost estimators).

The automatic derivation of heuristic functions for guiding the search provides what is probably the first fast and robust mechanism for carring out means-end analysis in complex domains.

7 Greedy Selection and Lookahead

Heuristic functions, as cost estimators, have also been found crucial for focusing the search in problems involving uncertainty and feedback where solutions are not 'paths' in the state space. Solutions to the various models can be all expressed in terms of control policies π that are *greedy* with respect to a given heuristic function h. A control policy π is a function mapping states $s \in S$ into actions $a \in A(s)$, and a policy π_h is greedy with respect to h iff π_h is the best policy assuming that the cost-to-go is given by h, i.e.

$$\pi_h(s) \;=\; \operatorname*{argmin}_{a \in A(s)} Q_h(a, s) \tag{1}$$

where $Q_h(a, s)$ is the expression of the cost-to-go whose actual form depends on the model; e.g., for non-deterministic models is $c(a, s) + \max_{s' \in F(a,s)} h(s')$, for MDPs $c(a, s) + \sum_{s' \in F(a,s)} P_a(s'|s)h(s')$, etc. In all cases, if the heuristic h is optimal; i.e., $h = V^*$, the greedy policy π_h is optimal as well [30, 4]. As mentioned above, the planner that won the last Probabilistic Planning Competition, used a greedy policy based on an heuristic function derived ignoring probabilistic information.

Often, if the heuristic estimator h is good, the greedy policy π_h based on it is good as well. Otherwise, there are two ways for improving the policy π_h without having to consider the entire state space: one is by *look ahead*, the other is by *learning,* and both involve *search*. Look-ahead is the strategy used in 2-player games like Chess that cannot be solved up to the terminal states; it is a variation of the greedy strategy π_h where the $Q_h(a, s)$ term is obtained not from the direct successors of s but from further descendants. The lookahead search is not exhaustive either, as values $h(s')$ of the tip

nodes are used to prune the set of nodes considered; a technique known as alpha-beta search [31]. The quality of the play depends on the search horizon and on the quality of the value function, which in this case, does not estimate cost but reward. In all the models, the greedy policy π_h is invariant to certain types of transformation in h; e.g $\pi_h = \pi_{h'}$ if $h' = \alpha h + \beta$ for constants α and β, $\alpha > 0$, so the value scale is not critical. Moreover, in Chess, any transformation of the heuristic function that preserves the relative ordering of the states, yields an equivalent policy, even if lookahead is used.

8 Learning

The second way to improve a greedy policy π_h is by adjusting the heuristic values during the search [32, 33, 34]. More precisely, after applying the greedy action

$$\operatorname*{argmin}_{a \in A(s)} Q_h(a, s)$$

in state s, the heuristic value $h(s)$ in s is updated to

$$h(s) := \min_{a \in A(s)} Q_h(a, s) \tag{2}$$

Interestingly, if h is admissible ($h \leq V^*$), and these updates are performed as the greedy policy π_h is simulated, the resulting algorithm exhibits two properties that distinguish it from standard methods: first, unlike a fixed greedy policy, it will never get trapped into a loop and will eventually get to the goal (if the goal is reachable from every state), and second, after repeated trials, the greedy policy π_h converges to an optimal policy, and the values $h(s)$ to the optimal values $V^*(s)$ (over the relevant states). This algorithm is called Real-Time Dynamic Programming (RTDP) in [33] as it combines a greedy, real-time action selecting mechanism, with the improvements brought about by the updates. Like heuristic search algorithms in AI but unlike standard DP methods, RTDP can solve large problems involving uncertainty, without having to consider the whole state space, provided that a good and admissible heuristic function h is used. Moreover, partial feedback can be accommodated as well, by performing the search in 'belief space'. GPT is a planner, that accepts description of problems involving stochastic actions and sensors, and computes optimal or approximate optimal policies using a refinement of these methods [10].

9 Inference

Many problems have a low polynomial complexity, and are easy for people to solve; e.g., the problem of collecting packages at various destinations in a city, and delivering them at some other destinations. This 'problem' is not even considered a problem by people as, unlike puzzles, can be solved (non-optimally) in a very straightforward way. Yet if the problem is fed to a planner by describing the actions of driving the truck from one location to another, picking up and loading the packages, and so on, the planner would tackle the problem in the same way it would tackle a puzzle: by means of *search*.

This search can often be done quite fast, yet like in Chess, this does not seem to be the way people solve such problems. Psychologists interested in problem solving, have focused on puzzles like Towers-of-Hanoi rather than on the simple problems that people solve every day. The work in planning however reveals that problems that are easy for people are not necessarily easy for a general automated problem solver. It may be argued that people solve these problems by using domain-specific knowledge, yet this pushes the problem one level up: how do people recognize when a problem falls in a domain, and how many domains are there? Recently we have addressed the related question of whether it is possible to solve a wide variety of 'simple' problems that are used as benchmarks in planning (including the famous Blocks World problems), by performing efficient (low polynomial) domain-independent inference and *no search.* To our surprise, we have found that this is possible [35]. We believe that there are a number of useful consequences to draw from this fact, given that most problems faced by real intelligent agents are not puzzles. In any case, inference and heuristic functions are two sides of the same coin: they both extract useful knowledge from a domain description and use it to focus the search, and if possible, to eliminate the search altogether.

10 SAT: Search and Inference in Logic

Logic has played a prominent role in AI as a basis for knowledge representation and programming languages. In recent years, logic restricted to propositional languages has become a powerful computational paradigm as well. A variety of problems can be encoded as SAT problems which are then fed and solved by powerful SAT solvers: programs that take a set of clauses (disjunctions of positive or negated atoms), and determine if the clauses are consistent, and if so, return a truth-valuation that satisfies all the clauses (a model). While the SAT problem is intractable, problems involving thousands of clauses and variables can now be solved [36]. Classical planning problems can be mapped into SAT by translating the problem descriptions into propositional logic, and fixing a planning horizon: if the theory is inconsistent, the problem has no solution within the horizon, else a plan can be read off the model [37]. For problems involving non-determinism, the SAT formulation yields only 'optimistic' plans, yet work is underway for reproducing the practical success of SAT in richer settings. Current SAT algorithms combine search and inference as well, and are complete. Some of the original algorithms, were based on local search [38], and were inspired by a neural-network constraint satisfaction engine [39].

11 Domain Compilation and Appraisal Circuits

Another recent development in logic relevant for action selection is *knowledge compilation* [40, 41]. In knowledge compilation, a formula is mapped into a logically equivalent formula of a certain form that makes certain class of operations more efficient. For example, while testing consistency of a formula is exponential in the size of the formula (in the worst case), formulas in d-DNNF can be tested in linear time (d-DNNF is a variation of 'Disjunctive Normal Form'; see [42, 43]). Moreover, for a formula T in

d-DNNF, it is possible, in linear-time as well (i.e., very efficiently) to check the consistency of $T + L$ for any set of literals L, get a model of $T + L$, and even count the number of such models. Of course, compiling a formula into d-DNNF is expensive, but this expense is worth if the result of the compilation is used many times. The idea of theory compilation has a number of applications in planning that are beginning to get explored. For example, Barret in [44], compiles planning theories with a fixed planning horizon n into d-DNNF, and shows that from the compiled theory *it is possible to obtain plans for arbitrary initial situations and goals, in linear-time with no search.* This is a very interesting idea that makes technical sense of the intuition that there are many logically equivalent representations, and yet some representations that are better adapted for a given task. We have recently explored a variation of Barret's idea that exploits another property of d-DNNF formulas T: the ability to efficiently compute not only models of T but also *best* models, 'best' defined in terms of a ranking over the individual (boolean) variables [45] that may encode penalties and rewards. In [46] it is shown, that the compilation of a suitable relaxation of T into d-DNNF renders a *circuit* that maps in linear-time, any given situation into a value that provides an appraisal of the situation. It does not take much to relate these appraisals with the role *emotions* in the selection of actions as postulated in a number of recent works; e.g., [47, 48, 49]. In the view that arises from domain compilation, however, emotions are prior to search, and they are not used for guiding the search or deliberation, nor are they the result of deliberation; rather they summarize expected reward or penalty (as when a deer sees a lion nearby). This view can account also for the way in which local preferences (ranks) are quickly assembled to provide an assessment of any situation (see "Feeling and Thinking: Preferences Need No Inferences" in [50]). In relation to the heuristic view of emotions, the notion of domain compilation provides an alternative and probably complementary view: in one case, emotion like an heuristic, guides the search for best reward, in the other, emotion stands for expected reward, which in the compiled representation is computed in linear-time (i.e., very quickly).

12 Summary

We have argued that humans encounter a huge variety of problems which they must solve using general methods. For general methods to work, however, they must be able to recognize and exploit structure. We have then reviewed some recent techniques from AI planning and problem solving that accomplish this, either focusing the search for solutions or bypassing the search altogether. These techniques include the automatic derivation of heuristic functions, the use of limited but effective forms of inference, and the compilation of domains, all of which enable a general problem solver to 'adapt' to the task at hand. We have also discussed briefly how these ideas relate to some biologically-motivated action selection models based on 'activation levels' and recent proposals linking emotions and search.

The area of planning and problem solving in Artificial Intelligence has come a long way, and it is probably time, following Marr's approach, to use the insights gained by the study of *what* is to be computed and *how* is to be computed, for gaining a better understanding of what real-brains actually compute when making plans and selecting

actions. Of course, there is a lot to be learned, and many other useful and necessary approaches to the problem, yet some of us hope that a good theory of AI planning and problem solving, as Newell, Simon, and others envisioned many years ago, will be an essential part of the global picture.

References

[1] Newell, A., Simon, H.: Elements of a theory of human problem solving. Psychology Review (1958)
[2] Newell, A., Simon, H.: GPS: a program that simulates human thought. In Feigenbaum, E., Feldman, J., eds.: Computers and Thought. McGraw Hill (1963) 279–293
[3] Smith, D.: Special issue on the 3rd international planning competition. JAIR **20** (2003)
[4] Bertsekas, D.: Dynamic Programming and Optimal Control, Vols 1 and 2. Athena Scientific (1995)
[5] Sutton, R., Barto, A.: Introduction to Reinforcement Learning. MIT Press (1998)
[6] Houston, A.I., McNamara, J.M.: A framework for the functional analysis of behaviour. Behavioral and Brain Sciences **11** (1988)
[7] Clark, C.: Modeling behavioral adaptations. Behavioral and Brain Sciences **14** (1991)
[8] Korf, R.: Finding optimal solutions to Rubik's cube using pattern databases. In: Proceedings of AAAI-98, AAAI Press / MIT Press (1998) 1202–1207
[9] Bonet, B., Geffner, H.: Planning as heuristic search. Artificial Intelligence **129** (2001) 5–33
[10] Bonet, B., Geffner, H.: Planning with incomplete information as heuristic search in belief space. In: Proc. of AIPS-2000, AAAI Press (2000) 52–61
[11] Glimcher, P.: Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics. MIT Press (2003)
[12] Tooby, J., Cosmides, L.: The psychological foundations of culture. In Barkow, J., Cosmides, L., Tooby, J., eds.: The Adapted Mind. Oxford (1992)
[13] Gigerenzer, G., Todd, P.: Simple Heuristics that Make Us Smart. Oxford (1999)
[14] Stanovich, K.: The Robot's Rebellion: Finding Meaning in the Age of Darwin. Chicago (2004)
[15] Brooks, R.: From earwigs to humans. Robotics and Autonomous Systems **20** (1997) 291–304
[16] McFarland, D., Bosser, T.: Intelligent behaviour in animals and robots. MIT Press (1993)
[17] Astrom, K.: Optimal control of markov decision processes with incomplete state estimation. J. Math. Anal. Appl. **10** (1965) 174–205
[18] Fikes, R., Nilsson, N.: STRIPS: A new approach to the application of theorem proving to problem solving. Artificial Intelligence **1** (1971) 27–120
[19] Penberthy, J., Weld, D.: Ucpop: A sound, complete, partiall order planner for adl. In: Proceedings KR'92. (1992)
[20] Blum, A., Furst, M.: Fast planning through planning graph analysis. In: Proceedings of IJCAI-95, Morgan Kaufmann (1995) 1636–1642
[21] Younes, H.L.S., Littman, M.L., Weissman, D., Asmuth, J.: The first probabilistic track of the international planning competition. Journal of AI Research **24** (2005) 851–887
[22] Hoffmann, J., Nebel, B.: The FF planning system: Fast plan generation through heuristic search. Journal of Artificial Intelligence Research **14** (2001) 253–302
[23] Hart, P., Nilsson, N., Raphael, B.: A formal basis for the heuristic determination of minimum cost paths. IEEE Trans. Syst. Sci. Cybern. **4** (1968) 100–107
[24] McDermott, D.: A heuristic estimator for means-ends analysis in planning. In: Proc. Third Int. Conf. on AI Planning Systems (AIPS-96). (1996)

[25] Bonet, B., Loerincs, G., Geffner, H.: A robust and fast action selection mechanism for planning. In: Proceedings of AAAI-97, MIT Press (1997) 714–719
[26] Cormen, T.H., Leiserson, C.E., Rivest, R.L.: Introduction to Algorithms. The MIT Press (1989)
[27] Pearl, J.: Heuristics. Addison Wesley (1983)
[28] Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall (1994)
[29] Maes, P.: Situated agents can have goals. Robotics and Autonomous Systems **6** (1990) 49–70
[30] Bellman, R.: Dynamic Programming. Princeton University Press (1957)
[31] Newell, A., Shaw, J.C., Simon, H.: Chess-playing programs and the problem of complexity. In Feigenbaum, E., Feldman, J., eds.: Computers and Thought. McGraw Hill (1963) 109–133
[32] Korf, R.: Real-time heuristic search. Artificial Intelligence **42** (1990) 189–211
[33] Barto, A., Bradtke, S., Singh, S.: Learning to act using real-time dynamic programming. Artificial Intelligence **72** (1995) 81–138
[34] Bonet, B., Geffner, H.: Learning in Depth-First Search: A unified approach to heuristic search in deterministic and non-deterministic settings. In: Proc. 16th Int. Conf. on Automated Planning and Scheduling (ICAPS-06), AAAI Press (2006)
[35] Vidal, V., Geffner, H.: Solving simple planning problems with more inference and no search. In van Beek, P., ed.: Proc. 11th Int. Conf. on Principles and Practice of Constraint Programming (CP 2005). Volume 3709 of Lecture Notes in Computer Science., Springer (2005) 682–696
[36] Kautz, H., Selman, B.: The state of SAT. Discrete and Applied Math (2005)
[37] Kautz, H., Selman, B.: Pushing the envelope: Planning, propositional logic, and stochastic search. In: Proceedings of AAAI-96, AAAI Press / MIT Press (1996) 1194–1201
[38] Selman, B., Levesque, H., Mitchell, D.: A new method for solving hard satisfiability problems. In: Proc. AAAI-92. (1992)
[39] Adorf, H.M., Johnston, M.D.: A discrete stochastic neural network algorithm for constraint satisfaction problems. In: Proc. the Int. Joint Conf. on Neural Networks. (1990)
[40] Selman, B., Kautz, H.: Knowledge compilation and theory approximation. Journal of the ACM **43** (1996) 193–224
[41] Darwiche, A., Marquis, P.: A knowledge compilation map. J. of AI Research **17** (2002) 229–264
[42] Darwiche, A.: Decomposable negation normal form. J. ACM **48** (2001) 608–647
[43] Darwiche, A.: On the tractable counting of theory models and its applications to belief revision and truth maintenance. J. of Applied Non-Classical Logics (2002)
[44] Barret, T.: From hybrid systems to universal plans via domain compilation. In: Proc. ICAPS-04. (2004)
[45] Darwiche, A., Marquis, P.: Compiling propositional weighted bases. Artificial Intelligence **157** (2004) 81–113
[46] Bonet, B., Geffner, H.: Heuristics for planning with penalties and rewards using compiled knowledge. In: Proc. 10th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR-06), AAAI Press (2006)
[47] Damasio, A.: Descartes' Error: Emotion, Reason, and the Human Brain. Quill (1995)
[48] Ketelaar, T., Todd, P.M.: Framing our thoughts: Evolutionary psychology's answer to the computational mind's dilemma. In III, H.H., ed.: Conceptual Challenges in Evolutionary Psychology. Kluwer (2001)
[49] Evans, D.: The search hypothesis of emotion. British J. Phil. Science **53** (2002) 497–509
[50] Zajonc, R.: The Selected Works of R.B. Zajonc. Wiley (2004)

The Neurodynamics of Visual Search

Gustavo Deco [1]
Institució Catalana de Recerca i Estudis Avançats (ICREA)
Universitat Pompeu Fabra
Dept. of Technology
Computational Neuroscience
Passeig de Circumval.lació, 8
08003 Barcelona, Spain

Josef Zihl
Ludwig-Maximilian University, Dept. of Psychology, Neuropsychology,
and Max Planck Institute of Psychiatry, Munich, Germany

Abstract. We review different functions in visual perception associated with attention and memory that have been integrated by a model based on the biased competition hypothesis. The model integrates, in a unifying form, the explanation of several existing types of experimental data obtained at different levels of investigation. At the microscopic level, single cell recordings are simulated. At the mesoscopic level of cortical areas, results of functional magnetic resonance imaging (fMRI) studies are reproduced. Finally, at the macroscopic level, the outcome of psychophysical experiments like visual search tasks are also described by the model. In particular, the model directly address how bottom-up and top-down processes interact in visual cognition, and show how some apparently serial processes re ect the operation of interacting parallel distributed systems. Attentional top-down bias guides the dynamics to focus attention at a given spatial location or on given features.

1 Introduction

To understand how the brain in general, and the visual brain in particular, works, it is necessary to combine different approaches, including neural computation. Neurophysiology at the single neuron level is needed because this is the level at which information

[1] Supported by ICREA

A. Fernández-Caballero & S. Miguel (Eds.): 50 Años de la Inteligencia Artificial, pp. 103-120, 2006.

is exchanged between the computing elements of the brain. Evidence from neuropsychology is needed to help understand what different parts of the system do and what each part is necessary for. Neuroimaging is useful to indicate where in the human brain different processes take place, and to show which functions can be dissociated from each other. Knowledge of the biophysical and synaptic properties of neurons is essential to understand how the computing elements of the brain work, and therefore what the building blocks of biologically realistic computational models should be. Knowledge of the anatomical and functional architecture of the cortex is needed to show what types of neuronal network actually perform the computation. And na lly the approach of neural computation is needed, as this is required to link together all the empirical evidence to produce an understanding of how the system actually works. This review utilizes evidence from some of these disciplines to develop an understanding of how vision is implemented by processing in the brain, focusing on visual attentional mechanisms.

2 Visual attentional mechanisms

As it is well known, because of the limited processing capacity of the visual system, attentional mechanisms are required in order to process information from a given scene. The dominant neurobiological hypothesis to account for attentional selection is that attention serves to enhance the responses of neurons representing stimuli at a single relevant location in the visual eld. This enhancement model is related to the metaphor for focal attention in terms of a spotlight (Treisman 1982, Tresisman 1988). This metaphor postulates a spotlight of attention which illuminates a portion of the eld of view where stimuli are processed in higher detail while the information outside the spotlight is l-tered out. According to this classical view, a relevant object in a cluttered scene is found by rapidly shifting the spotlight from one object in the scene to the next one, until the target is found. Therefore, according to this assumption the concept of attention is based on explicit serial mechanisms. The Feature Integration Theory of visual selective attention (Treisman and Gelade 1980 explains the outcome of numerous psychophysical experiments on visual search and offers an interpretation of the binding problem. In the feature integration theory, the rst preattentive process runs in parallel across the complete visual eld extracting single primitive features without integrating them. The second attentive stage corresponds to the serial specialized integration of information from a limited part of the eld at any one time. The main evidence for these two stages of attentional visual processing comes from psychophysical experiments using visual search tasks where subjects examine a display containing randomly positioned items in order to detect an a-priori de ned target.

There exists an alternative mechanism for selective attention, the *biased competition* model (Desimone and Duncan 1995,Duncan 1996,Duncan and Humphreys 1989. According to this model, the enhancement of attention on neuronal responses is understood in the context of competition among all of the stimuli in the visual eld. The *biased competition* hypothesis states that the multiple stimuli in the visual e ld activate populations of neurons that engage in competitive mechanisms. Attending to a stimulus at a particular location or with a particular feature biases this competition in favour of neurons that respond to the location or the features of the attended stimulus.

This attentional effect is produced by generating signals within areas outside the visual cortex which are then fed back to extrastriate areas, where they bias the competition such that when multiple stimuli appear in the visual eld, the cells representing the attended stimulus *win*, thereby suppressing cells representing distracting stimuli (Desimone and Duncan 1995,Duncan 1996,Duncan and Humphreys 1989). According to this line of work, attention appears as an emergent property of competitive interactions that work in parallel across the visual eld (for other alternative approaches see Heinke and Humphreys (2003))

Several neurophysiological experiments have been performed suggesting biased competition neural mechanisms that are consistent with such a hypothesis (Chelazzi, Miller, Duncan and Desimone 1993, Chelazzi 1998, Luck, Chelazzi, Hillyard and Desimone 1997,Reynolds,Chelazzi and Desimone 1999, Spitzer,Desimone and Moran 1988, Moran and Desimone 1985).

Further evidence comes from functional magnetic resonance imaging (fMRI) in humans (Kastner, De Weerd, Desimone and Ungerleider 1998,Kastner,Pinsk,De Weerd, Desomone and Ungerleider 1999). According to the biased competition hypothesis, these results show that when multiple stimuli are present simultaneously in the visual eld, their cortical representations within the object recognition pathway interact in a competitive, suppressive fashion, which is not the case when the stimuli are presented sequentially. It was also observed that directing attention to one of the stimuli counteracts the suppressive in uence of nearby stimuli.

3 The neurodynamical model

The overall systemic representation of the model is shown in Fig.1. The system is essentially composed of six modules (V1, V2-V4, IT, PP, v46, d46), structured such that they resemble the two known main visual paths of the mammalian visual cortex: the *what* and *where* paths (Deco and Zihl 1999,Hamker 1999,Deco and Zihl 2001,Rolls and Deco 2002). These six modules represent the minimum number of components to be taken into account within this complex system in order to describe the desired visual attention mechanism.

Information from the retino-geniculo-striate pathway enters the visual cortex through areas V1-V2 in the occipital lobe and proceeds into two processing streams. The occipital-temporal stream (*what* pathway) leads ventrally through V4 and IT (inferotemporal cortex) and is mainly concerned with object recognition, independently of position and scaling. The occipito-parietal stream (*where* pathway) leads dorsally into PP (posterior parietal cortex) and is concerned with the location of objects and the spatial relationships between objects. The model considers that feature attention biases intermodular competition between V4 and IT, whereas spatial attention biases intermodular competition between V1, V4 and PP.

The ventral stream consists of four modules: V1, V2-V4, IT, and a module v46 corresponding to the ventral area 46 of the prefrontal cortex, which maintains the short-term memory of the recognized object or generates the target object in a visual search task. The module V1 is concerned with the extraction of simple features, for example bars at different locations, orientation and size. This V1 module sends spatial and fea-

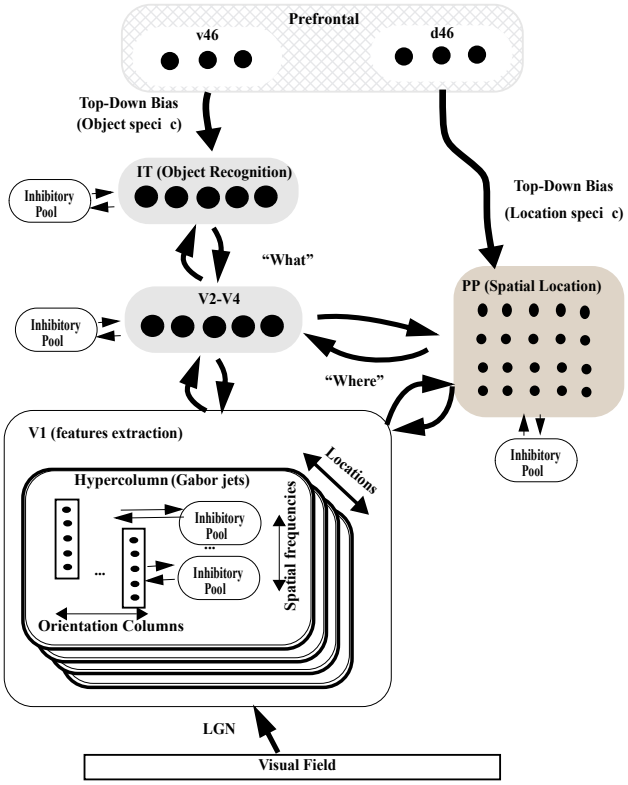


Fig. 1. Architecture of the neurodynamical approach. The system is essentially composed of six modules structured such that they resemble the two known main visual paths of the visual cortex.

ture information up to the dorsal and ventral streams. There is also one inhibitory pool interacting with the complex cells of all orientations at each scale. This inhibitory pool integrates information from all the excitatory pools within the module and feedbacks unspeci c inhibition uniformly to each of the excitatory pools. It mediates normalizing lateral inhibition or competitive interactions among the excitatory cell pools within the module. The module IT is concerned with the recognition of objects and consists of pools of neurons which are sensitive to the presence of a speci c object in the visual eld. It contains C pools, as the network is trained to search for, or recognize C particular objects. The V2-V4 module serves primarily to pool and channel the responses of V1 neurons to IT to achieve a limited degree of translation invariance. It also mediates a certain degree of localized competitive interaction between different targets.

The dorsal stream consists of three modules: V1, PP and d46. The module PP consists of pools coding the position of the stimuli, is responsible for mediating spatial attention modulation and for updating the spatial position of the attended object. The module d46 corresponds to the dorsal area 46 of the prefrontal cortex that maintains the short term spatial memory or generates the attentional bias for spatial location.

The prefrontal areas 46 (modules v46 and d46) are not explicitly modeled. Feedback connections between these areas provide the external top-down bias that speci es the task. The feedback connection from area v46 to the IT module speci es the target object in a visual search task. The feedback connection from area d46 to the PP module generates the bias to a targeted spatial location.

The system operates in two different modes: the learning mode and the recognition mode. During the learning mode, the synaptic connections between V4 and IT are trained by means of Hebbian learning during several presentations of a speci c object. During the recognition mode there are two possibilities of running the system. First, an object can be localised in a scene (visual search) by biasing the system with an external top-down component at the IT module which drives the competition in favour of the pool associated with the speci c object to be searched. Then, the intermodular attentional modulation V1-V4-IT will enhance the activity of the pools in V4 and V1 associated with the features of the speci c object to be searched for. Finally, the intermodular attentional modulation V4-PP and V1-PP will drive the competition in favour of the pool localising the speci c object. Second, an object can be identi ed (object recognition) at a speci c spatial location by biasing the system with an external top-down component at the PP module. This drives the competition in favour of the pool associated with the speci c location such that the intermodular attentional modulation V4-PP and V1-PP will favour the pools in V1 and V4 associated with the features of the object at that location. Intermodular attentional modulation V1-V4-IT will favour the pool that recognized the object at that location.

The neurons in the pools in **V1** have receptive elds performing a Gabor wavelet transform. Let us denote by I^{V1}_{kpql} the sensory input activity to a pool A^{V1}_{kpql} in V1 which is sensitive to a spatial frequency at octave k, to a preferred orientation de ned by the rotation index l, and to stimuli at the centre location speci ed by the indices pq. The sensory input activity to a pool in V1 (I^{V1}_{kpql}) is therefore de ned by the modulus of the corresponding Gabor wavelet transform of the image. Since in our numerical simulations the system needs only to learn a small number of objects (usually 2–4), we

temporarily did not include the V4 module for simplicity in some of the simulations. In fact, the large receptive fields of V2 and V4 can be approximately taken into account by including them in V1 pools with receptive fields corresponding to several octaves of the 2D-Gabor transform wavelets (i.e. not only the typical narrow receptive fields of V1 but also larger receptive fields are included in the modelled V1). The reduced system connects all cell assemblies in V1 with all cell assemblies in IT. The connections with the pools in the PP module are specified such that the modulation is Gaussian. Let us define in the PP module a pool A_{ij}^{PP} for each location ij in the visual field. The mutual (i.e. forward and back) connections between a pool A_{kpql}^{V1} in V1 and a pool A_{ij}^{PP} in PP are therefore defined by

$$w_{pqij} = A \exp\left\{-\frac{(i-p)^2 + (j-q)^2}{2S^2}\right\} - B \tag{1}$$

These connections mean that the V1 pool A_{kpql}^{V1} will have maximal amplitude when spatial attention is located at $i = p$ and $j = q$ in the visual field, i.e. when the pool A_{ij}^{PP} in PP is maximally activated and provides an inhibitory contribution $-B$ at the locations not being attended to. The V1–PP attentional modulation, in combination with the Hebbian learning that we will define later in this Section, generate translation-invariant recognition pools in the module IT.

Let us now define the neurodynamical equations that regulate the temporal evolution of the whole system.

The activity level of the input current in the **V1 module** is given by

$$\begin{aligned}\tau \frac{\partial A_{kpql}^{\mathrm{V1}}(t)}{\partial t} = &-A_{kpql}^{\mathrm{V1}} + \alpha \mathrm{F}(A_{kpql}^{\mathrm{V1}}(t)) - \beta \mathrm{F}(A_{k}^{\mathrm{I,V1}}(t)) + I_{kpql}^{\mathrm{V1}}(t) \\ &+ I_{pq}^{\mathrm{V1-PP}}(t) + I_{kpql}^{\mathrm{V1-IT}}(t) + I_0 + \nu\end{aligned} \tag{2}$$

where the attentional biasing coupling $I_{pq}^{\mathrm{V1-PP}}$ due to the intermodular 'where' connections with the pools in the parietal module PP is given by

$$I_{pq}^{\mathrm{V1-PP}} = \sum_{i,j} W_{pqij} \mathrm{F}(A_{ij}^{\mathrm{PP}}(t)) \tag{3}$$

and the attentional biasing term $I_{kpql}^{\mathrm{V1-IT}}$ due to the intermodular 'what' connections with the pools in the temporal module IT is defined by

$$I_{kpql}^{\mathrm{V1-IT}} = \sum_{c=1}^{C} w_{ckpql} \mathrm{F}(A_{c}^{\mathrm{IT}}(t)) \tag{4}$$

w_{ckpql} being the connection strength between the V1 pool A_{kpql}^{V1} and the IT pool A_{c}^{IT} corresponding to the coding of a specific object category c. We assume that the IT module has C pools corresponding to different object categories. For each spatial frequency level, a common inhibitory pool (designated with a superscript I) is defined. The current activity of these inhibitory pools obeys the following equations:

$$\tau_P \frac{\partial A_{k}^{\mathrm{I,V1}}(t)}{\partial t} = -A_{k}^{\mathrm{I,V1}}(t) + \gamma \sum_{p,q,l} \mathrm{F}(A_{kpql}^{\mathrm{V1}}(t)) - \delta \mathrm{F}(A_{k}^{\mathrm{I,V1}}(t)) \tag{5}$$

Similarly, the current activity of the excitatory pools in the **posterior parietal module PP** is given by

$$\tau \frac{\partial A_{ij}^{\mathrm{PP}}(t)}{\partial t} = -A_{ij}^{\mathrm{PP}} + \alpha \mathrm{F}(A_{ij}^{\mathrm{PP}}(t)) - \beta \mathrm{F}(A^{\mathrm{I,PP}}(t)) + I_{ij}^{\mathrm{PP-V1}}(t) + I_{ij}^{\mathrm{PP,A}} + I_0 + \nu \tag{6}$$

where $I_{ij}^{\mathrm{PP,A}}$ denotes an external attentional spatially-speci c top-down bias, and the intermodular attentional biasing $I_{ij}^{\mathrm{PP-V1}}$ through the connections with the pools in the module V1 is

$$I_{ij}^{\mathrm{PP-V1}} = \sum_{k,p,q,l} W_{pqij} \mathrm{F}(A_{kpql}^{\mathrm{V1}}(t)) \tag{7}$$

and the activity current of the common PP inhibitory pool evolves according to

$$\tau_P \frac{\partial A^{\mathrm{I,PP}}(t)}{\partial t} = -A^{\mathrm{I,PP}}(t) + \gamma \sum_{i,j} F(A_{ij}^{\mathrm{PP}}(t)) - \delta F(A^{\mathrm{I,PP}}(t)) \ . \tag{8}$$

The dynamics of the **inferotemporal module IT** is given by

$$\tau \frac{\partial A_c^{\mathrm{IT}}(t)}{\partial t} = -A_c^{\mathrm{IT}} + \alpha F(A_c^{\mathrm{IT}}(t)) - \beta F(A^{\mathrm{I,IT}}(t)) + I_c^{\mathrm{IT-V1}}(t) + I_c^{\mathrm{IT,A}} + I_0 + \nu \tag{9}$$

where $I_c^{\mathrm{IT,A}}$ denotes an external attentional object-speci c top-down bias, and the intermodular attentional biasing I_c^{IT-V1} between IT and V1 pools is

$$I_c^{\mathrm{IT-V1}} = \sum_{k,p,q,l} w_{ckpql} F(A_{kpql}^{\mathrm{V1}}(t)) \tag{10}$$

and the activity current of the common PP inhibitory pool evolves according to

$$\tau_P \frac{\partial A^{\mathrm{I,IT}}(t)}{\partial t} = -A^{\mathrm{I,IT}}(t) + \gamma \sum_{c} F(A_c^{\mathrm{IT}}(t)) - \delta F(A^{\mathrm{I,IT}}(t)) \ . \tag{11}$$

A more detailed explanation of the model can be found in (Rolls and Deco 2002,Deco,Pollatos and Zhil 2002).

4 Simulations of basic experimental ndings

4.1 Single cell experiments

Reynolds et al. (1999) rst examined the presence of competitive interactions in the absence of attentional effects, making the monkey attend to a location far outside the receptive eld of the neuron they were recorded from. They compared the ring activity response of the neuron when a single reference stimulus was located within the receptive eld with the response when a probe stimulus was added to the visual eld.

When the probe was added to the eld, the activity of the neuron was shifted towards the activity level that would have been evoked had the probe appeared alone. When the reference is an effective stimulus (high response) and the probe is an ineffective stimulus (low response) the ring activity is suppressed after adding the probe. In contrast, the response of the cell increased when an effective probe stimulus was added to an ineffective reference stimulus. Working within the neurodynamical model presented in Section 3, Deco and Lee (2002) and Corchs and Deco (2002)carried out numerical calculations to simulate these single cell experiments. Compared with the experimental results, the same qualitative behavior was observed for all experimental conditions analyzed. The competitive interactions in the absence of attention are due to the intramodular competitive dynamics at the level of V1, i.e. the suppressive and excitatory effects of the probe. The modulatory biasing corrections in the attended condition are caused by the intermodular interactions between V1 and PP pools, and PP pools and prefrontal top-down modulation.

4.2 fMRI experiments

The experimental studies of Kastner et al. (1998, 1999) show that when multiple stimuli are present simultaneously in the visual eld, their cortical representations within the object recognition pathway interact in a competitive, suppressive fashion. The authors also observed that directing attention to one of the stimuli counteracts the suppressive in uence of nearby stimuli. These experimental results were obtained by applying the functional magnetic resonance imaging (fMRI) technique in humans. In the rst experimental condition the authors tested the presence of suppressive interactions among stimuli presented simultaneously within the visual eld in the absence of directed attention, in the second experimental condition they investigated the in uence of spatially directed attention on the suppressive interactions, and in the third condition they analyzed the neural activity during directed attention but in the absence of visual stimulation. The authors observed that, because of the mutual suppression induced by competitively interacting stimuli, the fMRI signals were smaller during the simultaneous presentations than during the sequential presentations. In the second part of the experiment there were two main factors: presentation condition (sequential versus simultaneous) and directed attention condition (unattended versus attended). The average fMRI signals with attention increased more strongly for simultaneously presented stimuli than the corresponding signals for sequentially presented stimuli. Thus, the suppressive interactions were partially cancelled out by attention.

The dynamical evolution of activity at the cortical area level, as found in the behaviour of fMRI signals in experiments with humans, can be simulated in the framework of the neurodynamical model of Section 3 by integrating the pool activity in a given area over space and time. The integration over space yields an average activity of the considered brain area at a given time. With respect to the integration over time, it is performed in order to simulate the temporal resolution of fMRI experiments. Corchs and Deco (2002) simulated fMRI signals from V4 under the experimental conditions de ned by Kastner et al. As in the experiments, their simulations showed that the fMRI signals were smaller in magnitude during the simultaneous than during the sequential presentations in the unattended conditions because of the mutual suppression induced

by competitively interacting stimuli. On the other hand, the average fMRI signals with attention increased more strongly for simultaneously presented stimuli than the corresponding ones for sequentially presented stimuli. Thus, the suppressive interactions were partially cancelled out by attention.

4.3 Simulation of psychophysical experiments: visual search

We now concentrate on the macroscopic level of psychophysics. Evidence for different temporal behaviours of attention in visual processing comes from psychophysical experiments using visual search tasks where subjects scan a display containing randomly positioned items in order to detect an *a priori* de ned target. All other items in the display which are different from the target serve the role as distractors. The relevant variable tipically measured is search time as a function of the number of items in the display. Much work has been devoted on two kinds of search paradigm: feature search, and conjunction search. In a feature search task the target differs from the distractors in one single feature, e.g. only colour. In a conjunction search task the target is de ned by a conjunction of features and each distractor shares at least one of those features with the target. Conjunction search experiments show that search time increases linearly with the number of items in the display, implying a serial process. On the other hand, search times in a feature search can be independent of the number of items in the display. Deco and Zihl (2001) and Deco and Lee (2002) showed that the attentional architecture described in Section 3 performs search across the visual eld in parallel but, due to the different latencies of its dynamics, can show the two experimentally observed modes of visual attention, namely: serial focal attention, and the parallel spread of attention over space. The model demonstrates that neither explicit serial focal search nor saliency maps need to be assumed. The focus of attention is not provided to the system but only emerges after convergence of the dynamic behaviour of the neural networks.

To further elucidate these assumptions, Deco and Lee (20002) simulated the search of a letter located between other distractor letters. They de ned the task of searching for a letter ”E”. Figure 2 illustrates the basic observations concerning parallel and serial search.

The stimulus in Fig.2a contains shapes E and X. Because the elementary features in E and X are distinct, i.e. their component lines have different orientations, E pops out from X, and its location can be rapidly localized independently of the number of distracting X shapes in the display. On the other hand, the stimulus in Fig.2b contains the letters E (target) and F (distractors). Since both letters are composed of vertical and horizontal lines, there is no difference in elementary features to produce a preattentive pop-out, so they can be distinguished from each other only after their features are glued together by attention. It has been thought that because attention is serial, the time required to localize the target in such images increases linearly with the number of distractors of the display. The serial movement of the attentional spot light has been thought to be governed by a *saliency map* or *priority map* for registering the potentially interesting areas in the retinal input and directing a *gating* mechanism for selecting information for further processing. Does the linear increase in search time observed in visual search tests necessarily imply a serial search process, a saliency map or a gat-

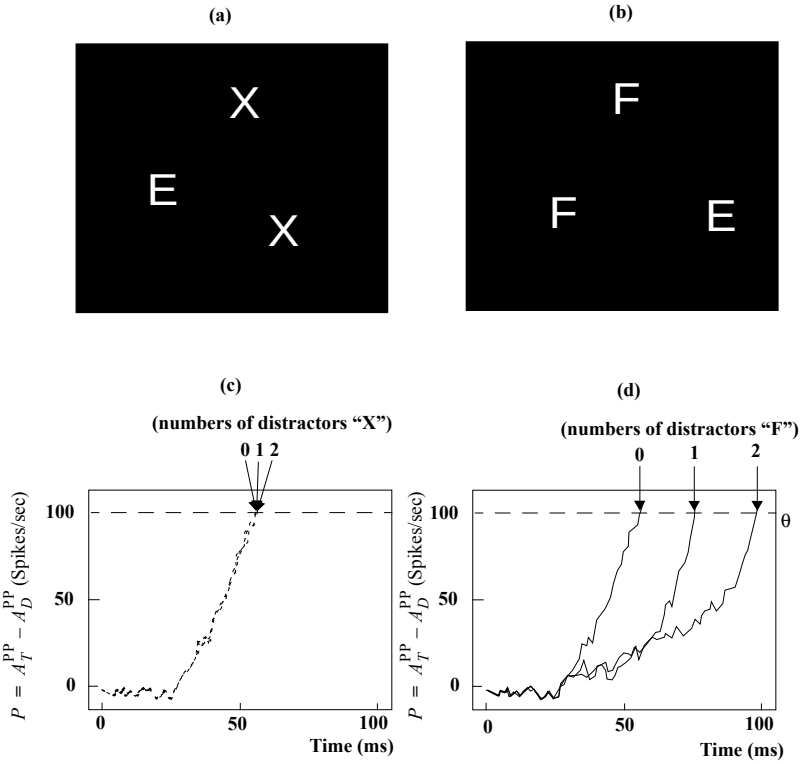


Fig. 2. (a) Parallel search example: an image that contains a target E in a eld of X distractors. Target E pops out from X. (b) Serial search example: an image that contains a target E in a eld of F distractors. The E and F can be distinguished from each other only after their features are bound by attention. (c-d) Simulation result of the network described in Section 3 performing visual search on images (a) and (b) respectively. The difference (polarization) between the maximum activity in the neuronal pool corresponding to the target locations and the maximum activity of all other neuronal pools in the dorsal PP module is plotted as a function of time.

ing mechanism? Could both the serial and parallel search phenomena be explained by a single parallel neurodynamical process without an additional serial control mechanism?

To investigate this issue, Deco and Lee (2002) presented the stimuli shown in Fig.2a and Fig.2b to the system described in Section 3, which has been trained to recognize X, E and F in a translation invariant manner. The system received a top-down bias for the 'E' pool in the ventral IT module, and then was presented with stimuli containing E in a variable number of X shapes, or E in a variable number of F shapes. Polarization (the difference between the maximal activity of the pools indicating the E location and that indicating the F location, i.e. $P = A_T^{PP} - A_D^{PP}$), is used as a measure to determine whether detection and localization of the target had been achieved or not. The authors found that for the E in X case, the time required for the polarization to reach a certain threshold in the dorsal PP module was almost identical whether the number of X shapes was equal to 0, 1 or 2, as shown in Fig.2c. On the other hand, when E and F were presented, the time required for polarization to reach threshold increased linearly with the number of distracting items. Although the system was running with the same parallel dynamics, it took an additional 25 msec for each additional distractor added to the stimulus as shown in Fig.2d. Therefore, the system works across the visual eld in parallel, but, due to the different dynamic latencies, resembles the two apparent different modes of visual attention: serial focal search and parallel search. The typical linear increase in search time with the display size is clearly obtained as the result of a slower convergency (latency) of the dynamics. In this case, the srong competition present in V1 and propagated to PP delays the convergency of the dynamics. The strong competition in the feature extraction module V1 is nally resolved by the feedback received from PP. In other words, stimulus similarity in the feature space is decided by competition mechanisms at the intramodular level of V1 and the intermodular level of V1-PP. The simulation results show that parallel search and serial search might not represent two essentially different independent stages as previously thought (Treisman and Gelade 1980). In the computational model described in Section 3, the two stages of processing (preattentive and attentive) involve the same mechanism and feature integration is accomplished dynamically by the interaction between the ventral IT module and the early V1 module. Feature integration is an emergent phenomenon due to interactive activation among the cortical areas, rather than a separate stage of visual processing, or involving a separate visual area.

Recently, we extended the neurodynamical model to account for the different slopes observed experimentally in complex conjunction visual search tasks (Deco and Zihl 2001,Deco et al. 2002,Rolls and Deco 2002). The authors assume that selective attention results from independent competition mechanisms operating within each feature dimension.

Quinlan and Humphreys (1987) analyzed feature search and three different kinds of conjunction search, namely: standard conjunction search and two kinds of triple conjunction with the target differing from all distractors in one or two features respectively. Let us de ne the different kinds of search tasks by using a pair of numbers m and n, where m is the number of distinguishing feature dimensions between target and distractors, and n is the number of features by which each distractor group differs from the target. Using this terminology, feature search corresponds to a 1,1-search; a standard

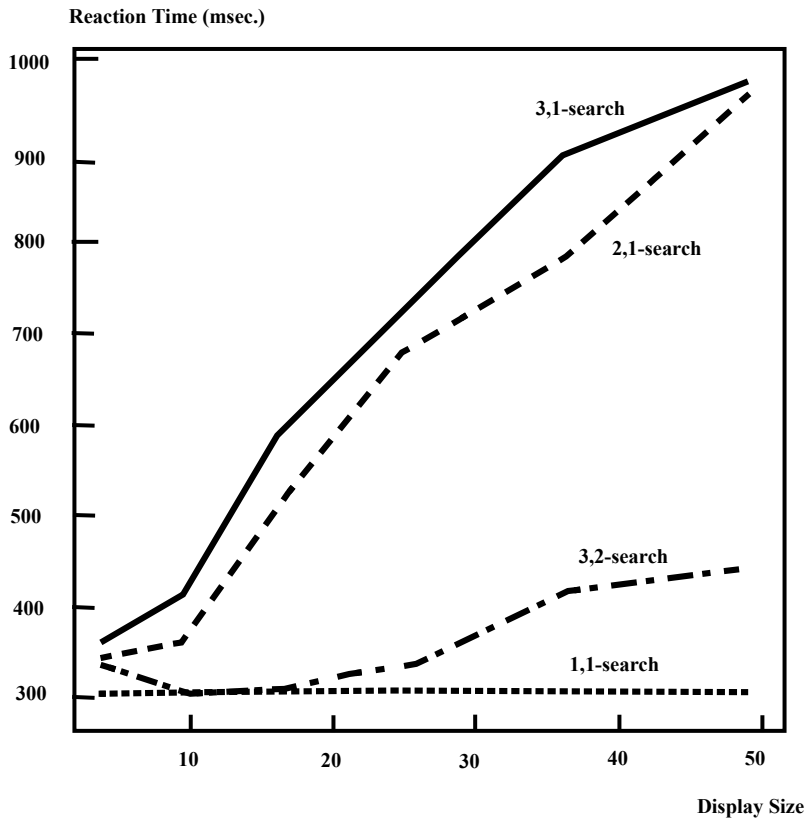


Fig. 3. Search times for feature and conjunction searches obtained utilizing the extended computational cortical model.

conjunction search corresponds to a 2,1-search; a triple conjunction search can be a 3,1 or a 3,2-search depending of whether the target difffers from all distractor groups by one or two features respectively. Quinlan and Humphreys (1987) showed that in feature search (1,1), the target is detected in parallel across the visual eld. They also show that the reaction time in both standard conjunction search and triple conjunction search conditions is a linear function of the display size. The slope of the function for the triple conjunction search task can be steeper or relatively at, depending upon whether the target differs from the distractors in one (3,1) or two features (3,2), respectively.

In Fig. 3, the computational results obtained by Deco and Zihl (2001) (see also (Deco et al. 2002)) for 1,1; 2,1; 3,1 and 3,2-searches are presented. The items are de- ned by three feature dimensions ($M = 3$, e.g. size, orientation and colour), each having two values ($N(m) = 2$ for $m = 1,2,3$, e.g. size: big/small, orientation: horizontal/vertical, colour: white/black). Figure 3 shows examples for each kind of search. For each display size, the experiment is repeated 100 times, each time with different randomly generated targets at random positions and randomly generated distractors. The mean value T of the 100 simulated search times is plotted as a function of the display size S. The slopes for all simulations are consistent with existing experimental results (Quinlan and Humphreys 1987).

5 Working memory and attention

In previous sections we have not explicitly modelled working memory, and we have used a mean- eld based neurodynamical approach in a set of networks with an externally applied attentional bias which affects the networks in a biassed competition scenario. In this section we summarize a model containing some of the working memory functions of the prefrontal cortex which provide the source of the biassed competition for the posterior perceptual areas in the parietal and temporal cortex. Moreover, we utilize a different approach to the dynamics in which each neuron in the network is modelled at the integrate-and- re level, so that we can produce spiking from the neurons in the network which can be directly compared with recordings from single neurons.

Working memory refers to an active system for maintaining and manipulating information in mind, held during a short period, usually of seconds (see Fuster (2000) for a more comprehensive de nition). The prefrontal cortex is involved in at least in some types of working memory, and neuronal activity in it continues during short term memory periods (Fuster 2000). Assad,Rainer and Miller (2000) investigated the functions of the prefrontal cortex in working memory by analyzing neuronal activity when a monkey performs two different working memory tasks using the same stimuli and responses. In a *conditional object-response (associative) task* with a delay, the monkey was shown one of two stimuli, and after a delay had to make either a rightward or leftward oculomotor saccade response depending on which stimulus was shown. In another experiment, recordings were made both during the object-response task and during a *delayed spatial response task*, in which the same stimuli were used, but the rule required was different, namely to respond towards the location where the stimulus had been shown (Assad et al. 2000). The main motivation for such studies was the fact that for real-world behavior, mapping between a stimulus and a response is typically

more complicated than a one-to-one mapping. The same stimulus can lead to different behavior depending on the context or the same behavior may be elicited by different cueing stimuli. In the performance of these tasks prefrontal cortex neurons were found that respond in the delay period to the stimulus object, the stimulus position ("sensory pools"), to combinations of response and stimulus object or position ("intermediate pools"), and to the response required (left or right) ("premotor pools").

The model is designed to help understand the underlying mechanisms that implement the working memory-related activity observed in neurons in the primate PFC in the context-dependent stimulus-response (associative) and delayed spatial response tasks investigated by (Assad et al. 2000). We build on the integrate-and- re attractor network treatment of Brunel and Wang (2001) and introduce a prefrontal cortex model with a hierarchically organized set of different attractor network pools (Deco and Roll 2003a). The hierarchical structure is organized within the general framework of the biased competition model of attention (Chelazzi 1998,Rolls and Deco 2002). The operation and parameters of the neurons in the integrate-and- r e model are similar to those of Brunel and Wang (2001), and are provided by Rolls and Deco (2002) and Deco and Rolls (2003).

Figure 4 shows schematically the synaptic structure assumed in the prefrontal cortical network. There are four excitatory populations or pools of neurons, namely: sensory, task or rule-speci c, premotor, and nonselective. The sensory pools encode information about objects, or spatial location. The premotor pools encode the motor response (in our case the leftward or rightward oculomotor saccade). The intermediate pools are task-speci c and perform the mapping between the sensory stimuli and the required motor response. The intermediate pools respond to combinations of the sensory stimuli and the response required, e.g. to object 1 requiring a left oculomotor saccade. The intermediate pools receive an external biasing input that re ects the current rule (e.g. on this trial when object 1 is shown make the left response after the delay period). The remaining excitatory neurons do not have speci c sensory, response or biasing inputs, and are in a nonselective pool. All the inhibitory neurons are clustered into a common inhibitory pool, so that there is global competition throughout the network.

Overall, the network has the architecture of a single attractor network with multiple activated populations or pools of neurons. These different pools engage in competitive interactions, are organized with some hierarchy imposed by the asymmetrically strong forward and backward connections, and receive biasing inputs to in uence the relative activity of the different pools, thus implementing attention-based or rule-based mapping from sensory inputs to motor outputs. This model thus shows how a rule or context input can in uence decision-making by biasing competition in a hierarchical network that thereby implements a e xible mapping from input stimuli to motor outputs (Rolls and Deco 2002,Deco and Rolls 2003b). The model also shows how the same network can implement a transient short term memory, and indeed how the competition required for the biased competition selection process can be implemented in an attractor network which itself requires inhibition implemented through the inhibitory neurons. Even more, the integrate-and re implementation of the network enables us to make explicit predictions of the effect of neuromodulation by manipulation of the dopamine level on the conditional object-response and delayed spatial response tasks.

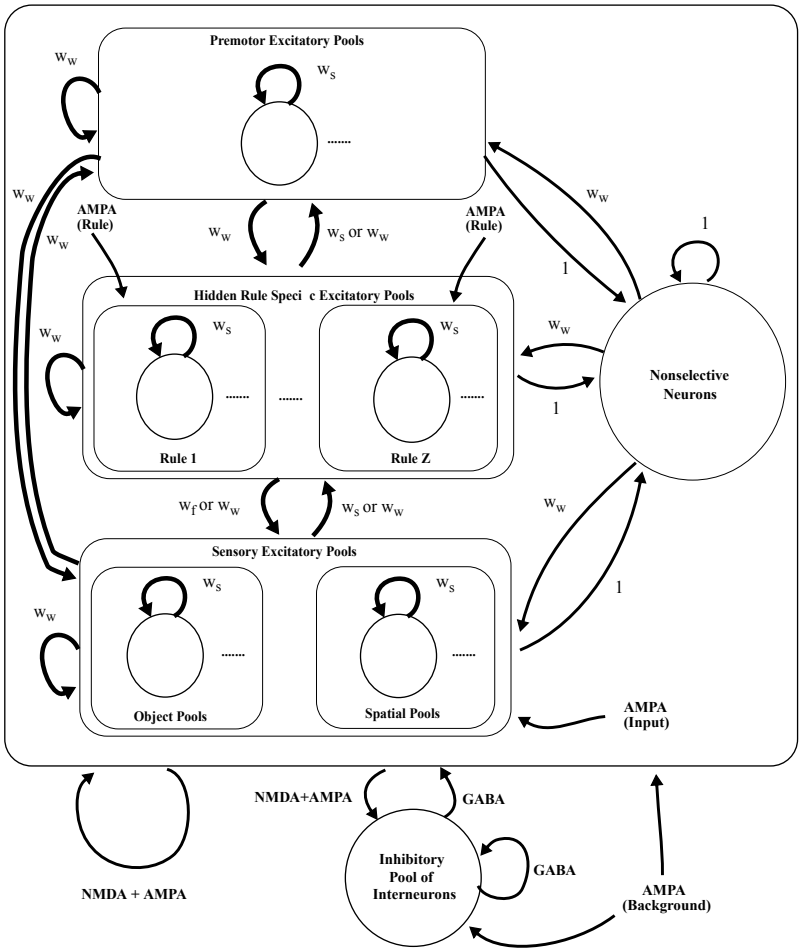


Fig. 4. Prefrontal cortical module.

A decrease in NMDA-related conductances produced by an increase in D2 receptor activation or a decrease in D1 receptor activation weakens and shortens persistent neuronal activity in transient short term memory periods. Additionally, we predict that the same pharmacological manipulations produce more response errors as a consequence of the more similar level of neuronal ring in the competing neuronal pools (Deco and Rolls 2003a).

6 Conclusions

The concept of biased competition has been used to account for object attention (Usher and Niebur 1996) and spatial attention in the ventral stream (Reynolds et al.1999). The neurodynamical model reviewed in this manuscript advances these ideas by bringing in the dorsal stream and the early visual areas to coordinate the organization of attention in a uni ed system (Deco and Zihl 2001,Rolls and Deco 2002). The two modes of attention emerge depending simply on whether a top-down bias is introduced to either the dorsal stream PP module or the ventral stream IT module. The spatial attention effect and competition interaction effect observed in the experiments of Moran and Desimone (1985) and Reynolds et al. (1999), can be accounted for by this model. It also shows and explains the dynamical competition and attention modulation effects found in attention experiments at the level of gross brain area activation as measured with fMRI (Kastner et al. 1999,Corchs and Deco 2002). In the context of visual search, the model shows that Treiman's feature integration (Treisman and Gelade 1980) can be implemented as an emergent phenomenon arising from the interaction between early visual cortical areas and the various extrastriate areas in the ventral and dorsal visual streams (Deco and Zihl 2001). The system works accross the visual eld in parallel, but, due to the different dynamic latencies, resembles the two apparent different modes of visual attention: serial focal search and parallel search.

In summary, computational neuroscience provides a useful mathematical framework for studying the mechanisms involved in brain function, like visual attentional mechanisms, that we have reviewed in the present work. The neurodynamical model here analyzed is based on evidence from functional, neurophysiological and psychological ndings. The simulations obtained with this theoretical model can successfully reproduce the experimental results of neurophysiological and fMRI experiments on spatial attention, as well as studies on serial and parallel search.

References

Asaad, W. F., Rainer, G. and Miller, E. K. (2000). Task-specific neural activity in the primate prefrontal cortex, *Journal of Neurophysiology* **84**: 451–459.

Brunel, N. and Wang, X. (2001). Effects of neuromodulation in a cortical networks model of object working memory dominated by recurrent inhibition, *Journal of Computational Neuroscience* **11**: 63–85.

Chelazzi, L. (1998). Serial attention mechanisms in visual search: a critical look at the evidence, *Psychological Research* **62**: 195–219.

Chelazzi, L., Miller, E., Duncan, J. and Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex, *Nature (London)* **363**: 345–347.

Corchs, S. and Deco, G. (2002). Large-scale neural model for visual attention: integration of experimental single cell and fMRI data, *Cerebral Cortex* **12**: 339–348.

Deco, G. and Lee, T. S. (2002). A unified model of spatial and object attention based on inter-cortical biased competition, *Neurocomputing* **44–46**: 775–781.

Deco, G. and Rolls, E. T. (2003a). Attention and working memory: A dynamical model of neuronal activity in the prefrontal cortex, *European Journal of Neuroscience* p. in press.

Deco, G. and Rolls, E. T. (2003b). Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex, *European Journal of Neuroscience, in press*.

Deco, G. and Zihl, J. (1999). A neural model of binding and selective attention for visual search, *in* D. Heinke, G. Humphreys and A. Olson (eds), *Connectionist Models in Cognitive Neuroscience – The 5th Neural Computation and Psychology Workshop*, Springer, Berlin, pp. 262–271.

Deco, G. and Zihl, J. (2001). Top-down selective visual attention: a neurodynamical approach, *Visual Cognition* **8**: 119–140.

Deco, G., Pollatos, O. and Zihl, J. (2002). The time course of selective visual attention:theory and experiments, *Vision Research* **42**: 2925 – 2945.

Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention, *Annual Review of Neuroscience* **18**: 193–222.

Duncan, J. (1996). Cooperating brain systems in selective perception and action, *in* T. Inui and J. L. McClelland (eds), *Attention and Performance XVI*, MIT Press, Cambridge, MA, pp. 549–578.

Duncan, J. and Humphreys, G. (1989). Visual search and stimulus similarity, *Psychological Review* **96**: 433–458.

Fuster, J. (2000). Executive frontal functions, *Experimental Brain Research* **133**: 66–70.

Hamker, F. (1999). The role of feedback connections in task-driven visual search, *in* D. Heinke, G. Humphreys and A. Olson (eds), *Connectionist Models in Cognitive Neuroscience – The 5th Neural Computation and Psychology Workshop*, Springer, Berlin, pp. 252–261.

Heinke, D. and Humphreys, G. (2003). Attention, spatial representation and visual neglect: Simulating emergent attention and spatial memory in the selective attention for identification model (saim), *Psychological Review* **110**: 29 – 87.

Kastner, S., De Weerd, P., Desimone, R. and Ungerleider, L. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI, *Science* **282**: 108–111.

Kastner, S., Pinsk, M., De Weerd, P., Desimone, R. and Ungerleider, L. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation, *Neuron* **22**: 751–761.

Luck, S., Chelazzi, L., Hillyard, S. and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex, *Journal of Neurophysiology* **77**: 24– 42.

Moran, J. and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex, *Science* **229**: 782–784.

Quinlan, P. T. and Humphreys, G. W. (1987). Visual search for targets defined by combination of color, shape, and size: An examination of the task constraints on feature and conjunction searches, *Perception and Psychophysics* **41**: 455–472.

Reynolds, J., Chelazzi, L. and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4, *Journal of Neuroscience* **19**: 1736–1753.

Rolls, E. T. and Deco, G. (2002). *Computational Neuroscience of Vision*, Oxford University Press, Oxford.

Spitzer, H., Desimone, R. and Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance, *Science* **240**: 338–340.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects, *Journal of Experimental Psychology: Human Perception and Performance* **8**: 194–214.

Treisman, A. (1988). Features and objects: The fourteenth Barlett memorial lecture, *The Quarterly Journal of Experimental Psychology* **40A**: 201–237.

Treisman, A. and Gelade, G. (1980). A feature-integration theory of attention, *Cognitive Psychology* **12**: 97–136.

Usher, M. and Niebur, E. (1996). Modelling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention, *Journal of Cognitive Neuroscience* **8**: 311–327.

Fundamentos de Neurobiología: aplicación a la neuroimagen

Ricardo Insausti, E. Artacho, M. del Mar Arroyo, X. Blaizot, A.M. Insausti, F. Mansilla*, P. Marcos-Rabal, A. Martínez-Marcos, A. Mohedano, M. Muñoz, P. Pro Sistiaga.

Laboratorio de Neuroanatomía Humana, Dpto. de Ciencias Médicas y Centro Regional de Investigaciones Biomédicas, Facultad de Medicina,
Universidad de Castilla-La Mancha, Albacete.
*Servicio de Radiología, Complejo Hospitalario Universitario de Albacete.

Resumen. La Neurobiología es una disciplina muy reciente, aunque su objeto, el estudio del sistema nervioso del hombre y de los animales, ha sido analizado desde la antigüedad clásica. En gran medida unido a la práctica neurológica clínica, aún continúa impartiéndose de modo fragmentario en gran medida. Desde la Anatomía, pasando por la Fisiología o Farmacología, el estudio del Sistema Nervioso se enfoca desde disciplinas más genéricas. La Neurobiología es una disciplina global que comprende todos aquellos aspectos, desde los moleculares (incluyendo los genéticos) hasta los más genéricos como pueden ser los estudios conductuales. El objetivo último de la Neurobiología podría definirse como aquel que busca el conocimiento de la organización del sistema nervioso y de sus manifestaciones (conducta). Dentro del amplio abanico de posibilidades que ofrece la Neurobiología, destacamos la Neuroimagen, en la que se trata de la visualización del sistema nervioso en el sujeto vivo, y que constituye una de las últimas especializaciones de la Neurobiología. El sistema nervioso, particularmente el humano, puede ser expuesto con gran detalle anatómico (p.e mediante la Resonancia Magnética estructural), mientras que la activación de los centros nerviosos implicados en actividades físicas o mentales puede ponerse de manifiesto con la Tomografía por Emisión de Positrones (PET) o la Resonancia Magnética Funcional (fMRI).

1 Introducción

La Neurobiología es un campo de la ciencia que se encarga de la investigación de los fundamentos de la anatomía (organización) y fisiología (funcionamiento) del Sistema Nervioso, tanto en humanos como en animales. La vertiente médica de la Neurobiología se encarga de la traslación bidireccional (o lo que es lo mismo, los problemas, modelos experimentales, y resultados propios de la Neurobiología son puestos en un contexto clínico, y, al mismo tiempo, las situaciones clínicas sirven de problema para el desarrollo de estudios experimentales) de los conocimientos neurobiológicos para

A. Fernández-Caballero & S. Miguel (Eds.): 50 Años de la Inteligencia Artificial, pp. 121-132, 2006.

abordar la patología del tejido nervioso y su diagnóstico clínico, así como su tratamiento (farmacología).

El tejido nervioso está formado por neuronas, células que responden a estímulos del mundo externo e interno con una respuesta eléctrica que es transmisible a otras neuronas, constituyendo redes neuronales y sistemas funcionales que son el soporte de la actividad conductual de los seres vivos, sea ésta sencilla o compleja. El conjunto de neuronas constituye el sustrato del tejido nervioso, y que precisan de otras células (glía) que apoyan la función de las neuronas en su actividad metabólica ,la cual es dependiente de un adecuado suministro sanguíneo (flujo circulatorio cerebral). La Neurobiología actual presenta un campo del saber unificado en un objetivo como es el conocimiento del sistema nervioso en salud (neurobiología básica, incluyendo neurobiología del desarrollo, y en particular la neurogenética, neuroanatomía, neurofisiología neuroquímica), neurobiología patológica (neuropatología, neurología, neurocirugía, neurofarmacología) y neurobiología de la conducta (psiconeurobiología) que engloba aspectos de la psicología y psiquiatría. La Neurobiología en sentido amplio se ha beneficiado de un tremendo impulso tecnológico en los últimos 15 años, de tal manera que se puede afirmar que su campo de aplicación, particularmente en el aspecto médico constituye uno de los pilares fundamentales en el planteamiento del gasto sanitario a nivel regional y nacional.

En esta breve introducción acerca de los fundamentos de Neurobiología, pretendemos exponer con un ejemplo de su cometido, en este caso la visualización de las estructuras del sistema nervioso, singularmente el cerebro, en el individuo vivo. Hasta ahora, la visualización de las lesiones cerebrales, p.e. tumores, quedaba indicada por el desplazamiento de los vasos sanguíneos que se visualizaban por la introducción de una sustancia radio-opaca (contraste) por vía arterial; la otra posibilidad tenía que ser el examen post-mortem, cuyo máximo exponente en Neurología se alcanzó en el S. XIX, sentando con ello las bases orgánicas de muchas enfermedades y síndromes cerebrales que hoy en día constituyen el cuerpo dela Neurología). La Neuroimagen, como subdisciplina de la Neurobiología más moderna presenta otra variante más dinámica y potencialmente más poderosa para contemplar la actividad nerviosa superior en el hombre. En efecto, la Neuroimagen cubre una gran cantidad de aspectos relacionados con la actividad funcional cerebral de actividades físicas y mentales que cubren campos que van desde la neuropsicología a la neurología clínica. La tecnología actual ofrece dos vertientes diferenciadas, aunque complementarias:

1) exposición en imágenes de estructuras cerebrales en situación estática, lo cual se consigue con la Resonancia Magnética estructural, la cual se realiza en situaciones normales o patológicas, y que ofrece la posibilidad de realizar análisis cuantitativos ,p.e. atrofia cerebral en determinadas enfermedades neurodegenerativas tales como la enfermedad de Alzheimer)

2) exposición de la actividad cerebral en situación dinámica, habitualmente mediante la captación de sustancias marcadas con isótopos radiactivos (PET), o la estimación dinámica de la variación del flujo cerebral (fMRI), las cuales nos revelan qué centros nerviosos presenten mayor actividad ante determinadas tareas experimentales que suponen una determinada actividad física o mental.

El avance del conocimiento de la base orgánica de la actividad cerebral, como ha quedado indicado anteriormente, ha estado restringido clásicamente a los estudios clínico-patológicos que dieron lugar al avance espectacular de la Neurología y Neurocirugía a lo largo del S. XIX y primera mitad del S. XX. En aquel tiempo, los neurólogos clínicos aprovechaban la enorme riqueza de síntomas y signos clínicos neurológicos que se asociaban al conocimiento anatómico de los centros y vías nerviosos, y que hoy en día continúan siendo una herramienta insustituible del diagnóstico clínico. De modo similar, un panorama equivalente existía en cuanto a los transtornos psiquiátricos en general, actualmente también diagnosticados mediante la entrevista clínica y tests psicológicos. Desde la antigüedad más clásica, el hombre se ha preguntado constantemente dónde y de qué manera el cerebro humano sostiene las características que nos definen como especie humana. Las respuestas han variado acorde con la época histórica, desde la unión de alma y cuerpo en la glándula pineal de Descartes en el S. XVII pasando por el mundo de la frenología en el S. XIX, la localización de determinadas funciones superiores ha venido dada por la observación de la pérdida de dichas funciones superiores ante lesiones cerebrales, bien de tipo vascular, tumoral o degenerativo, y la constatación de su asociación anatómica con regiones específicas del cerebro humano.

El desarrollo de la Neuroanatomía con la escuela de Cajal y otros contemporáneos hizo que el conocimiento de los centros y sistemas neurales cuya destrucción conllevaba la pérdida irreparable de funciones neurológicas y/o psicológicas, sentó las bases sobre las que se ha ido construyendo la moderna Neurobiología que abarca el estudio del sistema nervioso con todo el potencial que las diferentes disciplinas han ido desarrollando. Así tenemos desde el nivel iónico y molecular que nos identifica los mecanismos de la transmisión nerviosa (ver comunicación del Prof. Llinás en este curso) hasta los fenómenos eléctricos asociados a los estímulos sensoriales, la modulación de los mismos por diferentes fármacos, el estudio de la conducta en diferentes situaciones experimentales y su extrapolación a los trastornos conductuales en el hombre y el estudio de los sistemas neurales implicados.

Este cuerpo común de conocimientos que en muchos países constituye una licenciatura propia (no en el nuestro) es lo que hoy en día se estudia de un modo dinámico en el ser vivo (a diferencia de las épocas anteriores sólo se podía realizar mediante el examen post-mortem). En conjunto, esta nueva metodología de abordaje de cuestiones tanto clínicas (patologías del sistema nervioso central, incluyendo las psiquiátricas) como básicas (análisis de las funciones superiores del hombre) adopta una nueva forma de visualizarse en la Neuroimagen, término que engloba la visualización del cerebro humano, normal o patológico, bien desde un punto de vista estructural (resonancia magnética, MRI) o desde un punto de vista dinámico, bien mediante la captación de sustancias marcadas radiactivamente, como la tomografía por emisión de positrones (PET) o la resonancia magnética funcional (fMRI) que detecta cambios en el flujo cerebral asociado a un mayor metabolismo neuronal, y por tanto la puesta en marcha de una actividad neuronal focalizada, indicativa de la actividad de una serie de regiones cerebrales activadas en el desarrollo de una

de una serie de regiones cerebrales activadas en el desarrollo de una función específica (sistemas neurales).

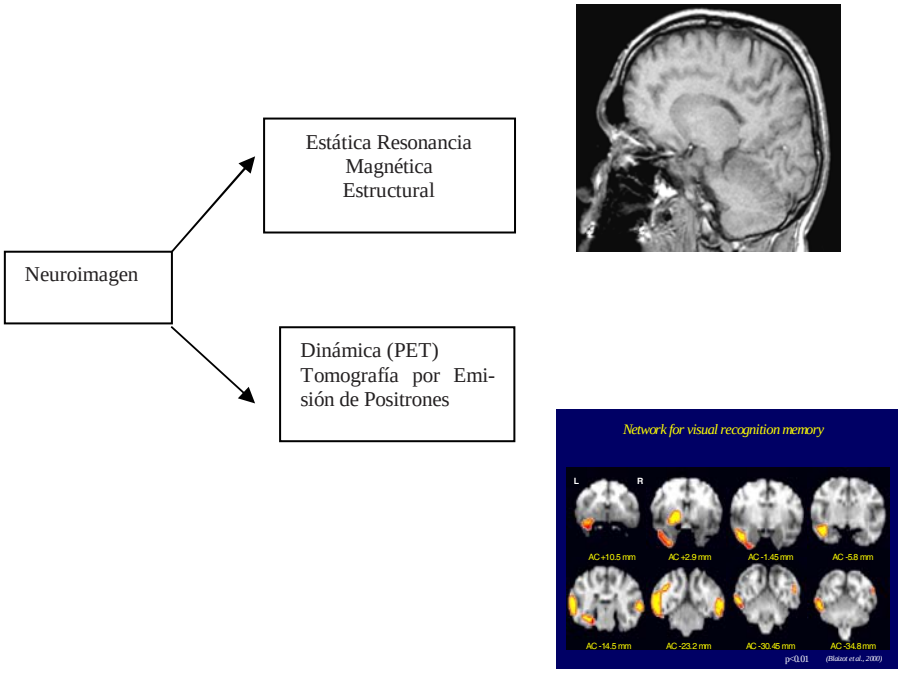


Figura 1: Técnicas fundamentales en Neuroimagen

2 Planteamiento.

En esta comunicación se va a presentar aspectos de la investigación que actualmente se está realizando en el Laboratorio de Neuroanatomía Humana de la Facultad de Medicina de la Universidad de Castilla-La Mancha en Albacete, conjuntamente con el Servicio de Radiología del Complejo Hospitalario Universitario de Albacete en el estudio de la determinación de estructuras cerebrales del hombre y del primate (macaco, babuino) relacionadas directamente con los mecanismos de formación, almacenamiento y recuperación de los recuerdos (memoria) como representativo de la Neuroimagen estructural, y que tienen una profunda implicación en alteraciones tan conocidas para la población general como la enfermedad de Alzheimer.

En una segunda parte, se van a exponer datos recientes de la exploración de la desintegración de la memoria visual en un modelo de lesión en babuinos, realizado en colaboración con el Centro Cicerón de la Universidad de Caen (Francia) con la Profesora Chantal Chavoix.

3 Estudios de correlación anatómica y de MRI.

Hemos analizado la estructura histológica del lóbulo temporal en el hombre y en el babuino y la hemos caracterizado según diferencias en la estructura de la corteza cerebral (citoarquitectura). El babuino, como una especia de primates superiores, es el único modelo de trabajo que ofrece una homología en su estructura con el lóbulo temporal humano, el cual la posee en un nivel mucho más elevado que ninguna otra especia animal. El análisis de las imágenes de MRI del cerebro del babuino permiten localizar con precisión estructuras tales como el hipocampo, corteza entorrinal, etc, todas ellas componentes esenciales de los sistemas neurales responsables de las funciones de memoria, y cuya lesión produce una amnesia profunda e irreversible, aunque tenga otros tipos de memoria (p.e. memoria reciente) intactas. Estas regiones, y en particular la corteza entorrinal, presenta las manifestaciones más precoces de la neurodegeneración que se aprecia en pacientes con enfermedad de Alzheimer. El hecho de que se localice precisamente en estas regiones la manifestación más acusada de la patología neurodegenerativa en la enfermedad de Alzheimer hace que su localización, delimitación de otras regiones circundantes y cuantificación volumétrica detallada, accesible "in vivo", permita desarrollar estrategias de seguimiento clínico a lo largo de los años, la evaluación de sus diferencias con respecto a sujetos de la misma edad, pero sin deterioro cognitivo, y la validación de terapéuticas. No memos importante es el desarrollo de modelos experimentales, particularmente en especies próximas al hombre, como son los primates, y en particular macacos y babuinos, que repliquen síntomas clínicos de la enfermedad de Alzheimer, así como terapias conducentes al tratamiento de esta enfermedad, cuya característica más esencial es la incapacidad de formar nuevas experiencias de memoria autobiográfica.

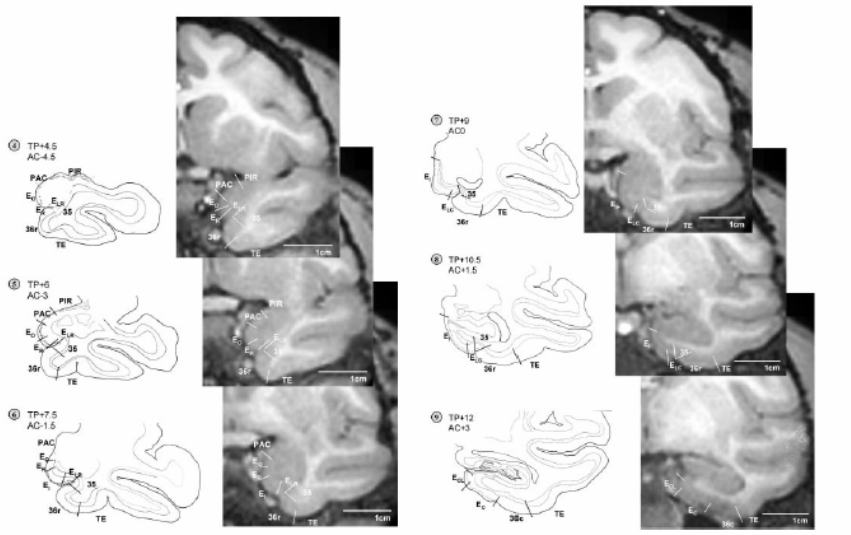


Figura 2. Estructuras anatómicas de los sistemas de memoria en el babuino (MRI), Blaizot y cols., 2004)

4 Estudios de la atrofia del lóbulo temporal en el envejecimiento normal y en la enfermedad de Alzheimer

Estamos estudiando la localización por MRI de una región del lóbulo temporal en sujetos normales a lo largo delas últimas décadas de la vida, y en una población afectada por enfermedad de Alzheimer y otras demencias. Este estudio se está realizando en colaboración con los Servicios de Geriatría (Dr. Abizanda) y de Radiología (Dr. Mansilla) del Complejo Hospitalario Universitario de Albacete. Esta región es instrumental en el proceso de elaboración de estímulos visuales y auditivos complejos, que permiten la localización espacial de los estímulos (dónde se encuentran en el espacio personal) así como de la interpretación de imágenes complejas (p.e. caras de otras personas) y de su interpretación personal. Esta región recibe el nombre de región parahipocámpica, la cual posee un poderoso influjo sobre la corteza entorrinal, y en consecuencia sobre el hipocampo. Los estudios detallados de su caracterización histológica y su correlación con el primate, ofrecen una posibilidad de diagnóstico precoz de una demencia tipo Alzheimer, ya que el tejido nervioso que la constituye sufre una atrofia perfectamente identificable y cuantificable.

Previamente, hemos desarrollado técnicas que permiten identificar con gran aproximación la delimitación de regiones más rostrales en el lóbulo temporal humano, observándose una correlación estrecha entre el grado de demencia y el volumen de estas regiones. El fundamento último es el hecho conocido de la muerte neuronal que acompaña a estos procesos, el cual se compaña inevitablemente de la pérdida de sustancia cerebral (neuronas más fibras nerviosas) en una magnitud que permite su detección y cuantificación, siendo un procedimiento diagnóstico de amplia utilización.

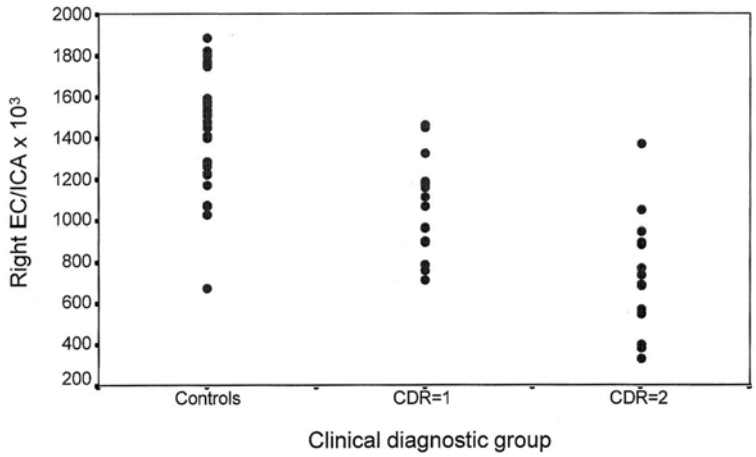


Figura 3: Reducción del volumen de la corteza entorrinal en grados sucesivos de demencia en la enfermedad de Alzheimer (de Juottonen, Insausti et al., 1998)

En la actualidad, nos encontramos refinando la localización de los componentes de la región parahipocámpica en imágenes de Resonancia Magnética obtenidos en cerebros postmortem, los cuales son examinados neuroanatómicamente para la exacta delimitación de las regiones constituyentes y de su atrofia diferencial en casos de demencia.

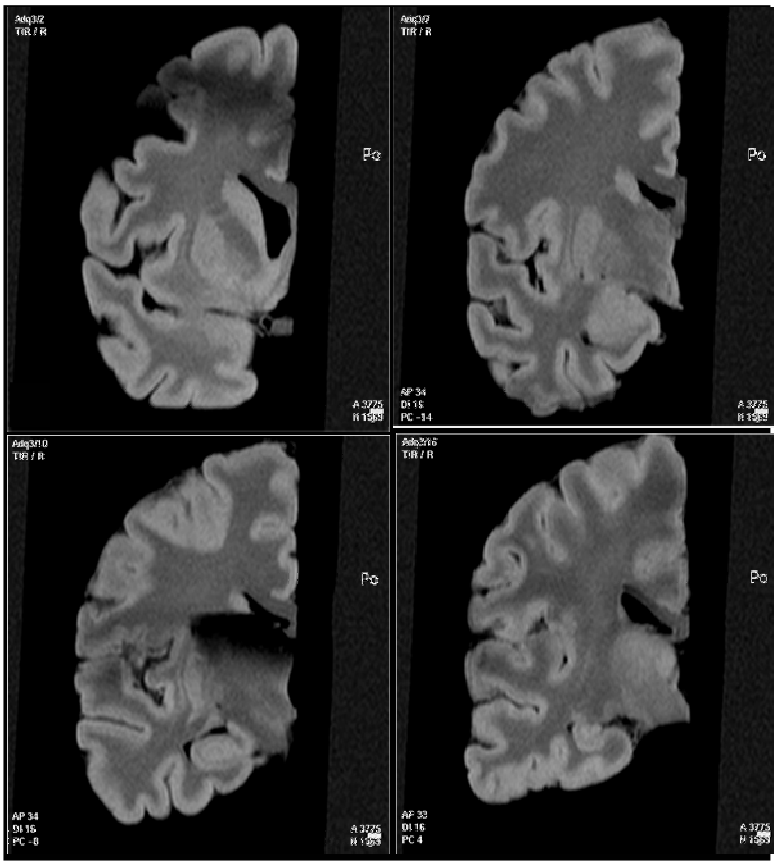


Figura 4: Imágenes de Resonancia Magnética de un hemisferio humano postmortem, fijado en formol. Los recuadros blancos indican la región parahipocámpica objeto del estudio

5 Estudios dinámicos de pérdida de memoria en el babuino

Presentamos unas imágenes que muestran la activación de la zona más anterior de la región parahipocámpica en el lóbulo temporal del babuino, antes y después de la lesión de uno de los componentes de los sistemas neurales de la memoria (corteza

perirrinal). Estos experimentos ponen en evidencia que el cerebro se reorganiza, de tal modo que circuitos y estructuras neuronales que no respondían a las tareas de memoria (las imágenes de PET se obtienen inmediatamente después de que el babuino realice el test de reconocimiento visual).

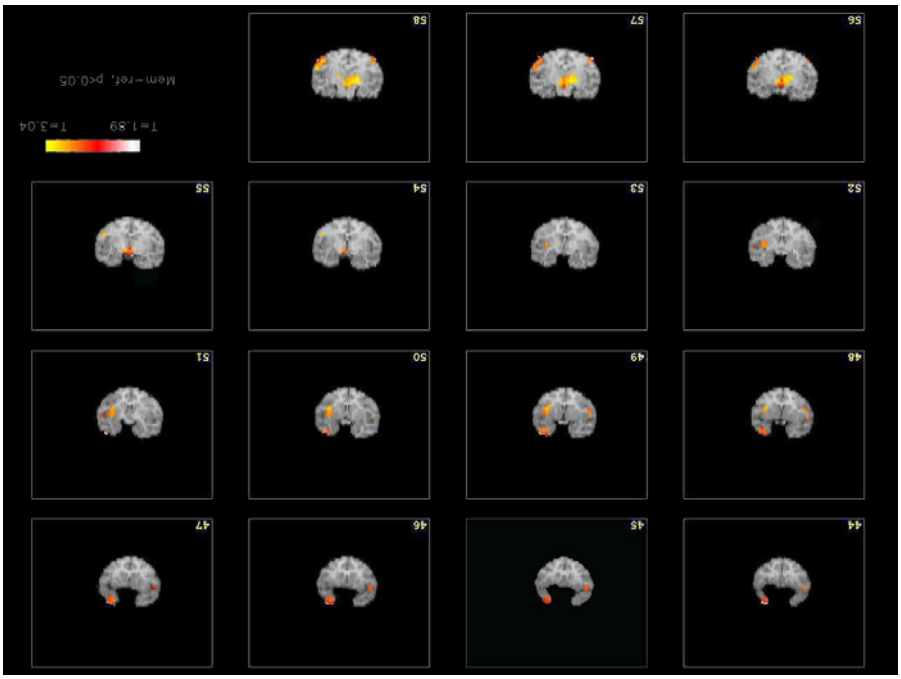


Figura 5: Activación de la región parahipocámpica (flechas) en un babuino en situación preoperatoria, en la que destacan las zonas activadas en el test de memoria visual. Las imágenes inferiores corresponden a los niveles más anteriores de la región parahipocámpica. Otras zonas activadas responden a la activación concomitante desarrollada en el test, y que no difieren de la situación basal.

La lesión de la región parahipocámpica conduce a una eliminación de la actividad en el animal sometido al mismo test, y revela la activación de zonas anteriormente silentes.

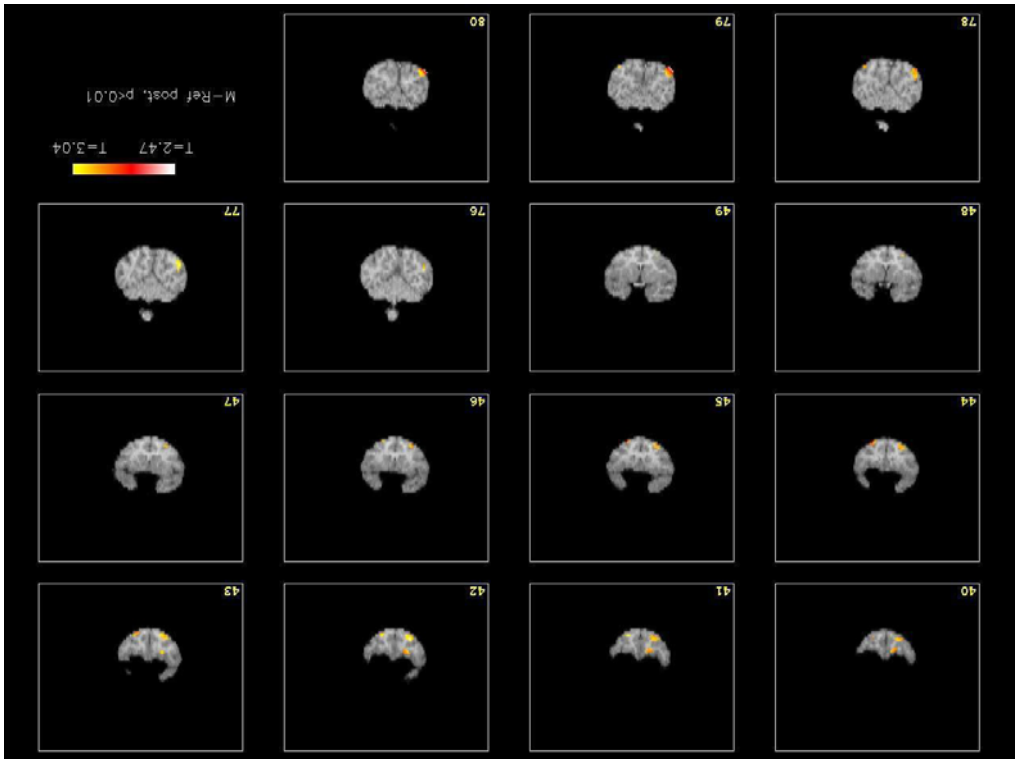


Figura 6: Desaparición de la activación cerebral de la región parahipocámpica tras su lesión por ácido iboténico , situación postoperatoria (tomado de G.Rauchs, X. Blaizot, C. Giffard, J.C. Baron, R. Insausti and C. Chavoix , 2006)

5 Otros estudios funcionales.

A continuación se presentan unos ejemplos de actividades que se realizan actualmente en muchos centros de investigación, aunque estas técnicas no son asequibles en nuestra región.

Los estudios varían en función de las tareas a realizar, pero en conjunto ilustran lo que en estos momentos la Neuroimagen ofrece no sólo en cuanto a posibilidades diagnósticas, sino también en cuanto a la comprensión de los fenómenos biológicos que subyacen a la estructura de la personalidad.

Entre la enorme variedad y abundancia de estudios ofrecemos ejemplos de:

- actividad visual (representación mental de imágenes),

- representación del dolor y del sufrimento
- representación de la actividad musical
- actividad cerebral en la depresión
- actividad cerebral en el envejecimiento y enfermedad de Alzheimer
- actividad cerebral en el pensamiento

Aunque tan sólo se trata de una somera exposición de las diferentes condiciones y situaciones que hoy en día se están aplicando en muchos centros de investigación, no cabe duda de que en un futuro inmediato se va a tratar de obtener información objetiva que nos va a ayudar en la comprensión de nosotros mismos, como seres biológicos con manifestaciones psíquicas que nos diferencian de los animales, aunque está por determinar en qué medida, y cómo este conocimiento nos va a afectar como individuos y como especie.

Agradecimientos

Estos estudios se realizan con ayudas a la investigación del Ministerio de Educación y Ciencia, y de la Junta de Comunidades de Castilla-La Mancha a través de las Consejerías de Educación y de Sanidad.

Bibliografía

Blaizot X, Martinez-Marcos A, Arroyo-Jiménez MM, Marcos P, Artacho E, Muñoz M, Chavoix C. Insausti R. The parahippocampal gyrus in the baboon: anatomical, cytoarchitectonic and MRI studiesCerebral Cortex. 14: 231-246 (2004.).

Insausti, K. Juottonen, H. Soininen, A. Insausti, K. Partanen, P. Vainio, M.P. Laakso, and a. Pitkanen MRI-based volumetric analyses of the human entorhinal, perirhinal and temporopolar cortices. American Journal of Neuroradiology 16: 659-671 (1998)

K. Juottonen, M.P. Laakso, R. Insausti, M. Lehtovirta, A. Pitkänen, K. Partanen and H. Soininen. (1998) Volumes of the entorhinal and perirhinal cortices in Alzheimer's disease. Neurobiology of Aging 19: 15-22

Kandel E. R, JH Schwartz, and T. Jessell Principles of Neural Sience, 4th edition, McGraw-Hill, New York, 2000.

Toga A.W. and J.C. Mazziotta, Eds. Brain Mapping. The Systems. Academic Press, San Diego, 2000. [Shannon & McCarthy eds., 56] Shannon, C.E., McCarthy, J. (Eds.): Automata Studies. Princeton University Press, N. Jersey (1956).

La inteligencia desde el punto de vista de la Psicología

José Miguel Latorre Postigo

Dpto. Psicología
Facultad de Medicina, Albacete.
Universidad de Castilla-La Mancha

PRÓLOGO

Hay una película de Stanley Kubrick (1968), basada en la obra de Artur C. Clarke, titulada "***2.001, Una Odisea Espacial***" que, aunque sujeta a tantas interpretaciones como personas la vean, a mí me parece que habla sobre la evolución de la inteligencia y del ser humano[1]. Un monolito que representa la inteligencia es plantado al lado de un grupo de simios (*Australopitecus Afarensis*) hace 4 millones de años. Impulsados por sus características básicas (miedo, curiosidad, valentía) hacen un descubrimiento revolucionario: la herramienta (en forma de hueso). Con la excitación del descubrimiento lanzan el hueso-herramienta al aire y esa imagen nos lleva hasta la estación espacial que representa la herramienta desarrollada por el hombre actual (civilizado, racional y científico). Pero en el espacio el hombre pierde el control de sus herramientas, que empiezan a tomar forma humana. El ordenador HAL 9000, cerebro y sistema nervioso de la estación espacial *Discovery*, controla todo lo que ocurre a su alrededor. ¿Para qué sirven los hombres? Son meros encargados de mantenimiento, la máquina puede prescindir de ellos. Al final de la película, llega el hombre inteligencia pura: el niño de las estrellas. ¿Es la inteligencia una propiedad independiente del ser humano? Aunque no podamos responder a esta pregunta, sí podemos decir que la inteligencia, entendida como la capacidad de solucionar problemas cada vez más complicados, ha ido evolucionando de forma paralela a la evolución de la especie humana. El árbol de la evolución del homo sapiens (figura 1) ha venido marcado por la evolución del cerebro y pon ende de la inteligencia. El volumen cerebral ha aumentado 1000 cm^3 en 4 millones de años, desde los 400 en el afarensis hasta los 1400 del sapiens. Además, la mayor especialización del neocortex ha ido acompañando al desarrollo de las habilidades inteligentes, adaptándose paulatinamente desde la búsqueda efectiva de alimentos hasta la búsqueda de soluciones a los complejos problemas sociales y tecnológicos (Geary, 2005).

[1] En palabras del propio Kubrick: "No es un mensaje que yo haya tratado de convertir en palabras. '2001' es una experiencia visual: de dos horas y 19 minutos de película, sólo hay un poco de menos de 40 minutos de diálogo. Traté de crear una experiencia visual que trascendiera las limitaciones del lenguaje y penetrara directamente en el subconsciente con su carga emotiva y filosófica. Como diría McLuhan, en '2001' el mensaje es el medio. Quise que la película fuera una experiencia intensamente subjetiva que llevara al espectador a un nivel interno de conciencia, como lo hace la música. "Explicar" una sinfonía de Beethoven sería castrarla, levantando una barrera artificial entre la concepción y la apreciación.

A. Fernández-Caballero & S. Miguel (Eds.): 50 Años de la Inteligencia Artificial, pp. 131-154, 2006.

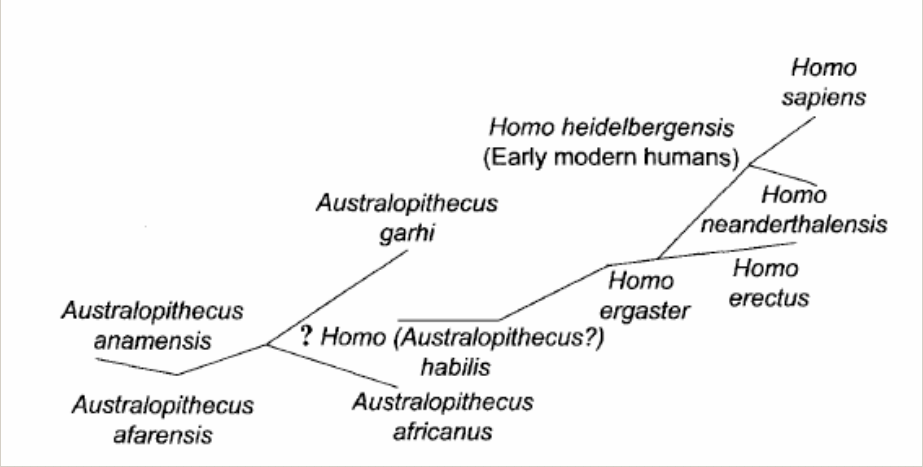


Figura 1: El árbol de la evolución de la familia del homo sapiens (Geary, 2005)

Además, la inteligencia sigue aumentando. Lo que se conoce como el *efecto Flynn* es el hecho de que año tras año el promedio del CI obtenido en los test aumenta en la mayoría de los países del mundo (Dickens y Flynn, 2001). Concretamente, se produce un aumento de tres puntos de CI por década. Se han formulado diversas explicaciones como la mejora en la nutrición, la disminución del tamaño de las familias, la mejora de la educación, el incremento de la complejidad del medio o la heterocigosidad (Mingroni, 2004).

INTRODUCCIÓN

Inteligencia, test de inteligencia, cociente de inteligencia, a menudo nos encontramos con estos conceptos mágicos, que utiliza todo el mundo. Describimos una persona como más o menos inteligente, al discutir problemas se buscan soluciones inteligentes o incluso se utilizan expresiones relacionadas con la inteligencia para insultar a otro. También se puede llegar a decir que inteligencia es lo que miden los test de inteligencia, y es que el concepto de inteligencia en su origen está muy unido a la utilización de los test. Sin embargo, el concepto de inteligencia que se esconde tras el test de inteligencia es demasiado estricto y a su vez controvertido. La utilización de test de inteligencia o de aptitud está plenamente extendida, incluso con tendencia a aumentar, en la orientación de empresas, en el ejército, o en las admisiones en algunas escuelas o centros de formación, en la orientación profesional de las oficinas de empleo y también como ayuda psicológica o psiquiátrica. En la actualidad, sin embargo, han aparecido diferentes teorías que plantean la existencia de *inteligencias múltiples*. Se habla de *inteligencia emocional*, *inteligencia social* o *inteligencia musical*, entre otras. Además de todo lo anterior, la investigación sobre la inteligencia rebasa el campo de la Psicología y, en el contexto de la *ciencia cognitiva*, abarca ciencias como la Filosofía, la Computación o la Neurofisiología, entre otras. En este trabajo nos vamos a centrar en el estudio de la inteligencia, tal y como se entiende desde las distintas áreas de la Psicología.

DEFINICIÓN DE INTELIGENCIA

Inteligencia, de acuerdo con el diccionario de la *Real Academia Española de la Lengua* (vigésima segunda edición), es la "capacidad de entender o comprender", la "capacidad de resolver problemas, el "conocimiento, comprensión o acto de entender", o la "habilidad, destreza o experiencia". Como definición expresa algunas de las facetas de la naturaleza de la inteligencia, pero no necesariamente lo que los psicólogos consideran central.

Una forma sencilla de entender el concepto de inteligencia es hacer que expertos o profanos lo definan. Sternberg (1988) parte de este supuesto y expone, en primer lugar, la opinión de los expertos respecto a la naturaleza de la inteligencia. Estos expertos (todos ellos de EEUU) fueron convocados por los directores del *Journal of Educational Psychology* en un simposio titulado "La inteligencia y su medida", en 1921, invitándoles a escribir sobre diversos aspectos en relación con la inteligencia. Más recientemente, en un simposio celebrado en 1986, se han tratado las mismas cuestiones planteadas en 1921, pero abordando el estudio de la inteligencia desde la *ciencia cognitiva* (Sánchez-Cánovas, 1995).

Los elementos centrales de la definición de inteligencia, de acuerdo con los expertos, son (a) la capacidad de adaptación destinada e enfrentarse con eficacia a las demandas externas, (b) los procesos elementales como atención y percepción, (c) los procesos complejos de razonamiento abstracto, representación mental, solución de problemas, toma de decisiones, (d) la habilidad para aprender, y (e) la conducta eficaz en respuesta a los problemas situacionales (Sternberg, 2003).

En el simposio celebrado en 1986 se abordó la inteligencia desde la perspectiva de la *ciencia cognitiva*[2], se hizo más hincapié en que todos los sistemas inteligentes comparten la *capacidad para manejar símbolos*. Pero, ¿cuáles son las propiedades fundamentales de los sistemas inteligentes? Los investigadores en ciencia cognitiva están de acuerdo en que lo común y básico es la *computación*. Aun cuando algunos autores, como Norman, admiten que los sistemas inteligentes biológicos no son sólo computación, todos están de acuerdo que una cualidad clave de los sistemas inteligentes es su *capacidad para el manejo de símbolos.* Jonson-Laird (1990) subraya esta idea cuando afirma: "Un principio fundamental de la ciencia cognitiva es que *la mente es un sistema simbólico*".

---

[2] Para Norman (1987), la **ciencia cognitiva** es una disciplina creada a partir de una convergencia de intereses entre los que se persiguen el estudio de la cognición desde diferentes puntos de vista. El aspecto crítico de la ciencia cognitiva es la búsqueda de la comprensión de la cognición, sea ésta real o abstracta, humana o mecánica. Su meta es comprender los principios de la conducta inteligente. Su esperanza es que ello nos permita una mejor comprensión de la mente humana, de la enseñanza y del aprendizaje, de las habilidades mentales y el desarrollo de aparatos inteligentes que puedan aumentar las capacidades humanas de manera importante y constructiva. La ciencia cognitiva se ha desarrollado, en cierto sentido, a través de la conjunción de la investigación de la inteligencia artificial con otras ciencias tales como la psicología cognitiva, la neurofisiología y la lingüística. El enfoque de la ciencia cognitiva se basa en la metáfora de la cognición del procesamiento de la información, y un requisito general es que modelos y teorías deben ser computacionales de modo que sean verificables por medio de la simulación a través del computador.

Algunos expertos, sin embargo, han definido la inteligencia de manera operacional, diciendo simplemente que es "lo que miden los test de inteligencia". Esta definición, propuesta por primera vez por Edwin Boeing en 1923 es en última instancia circular, ya que para poder desarrollar los test es necesario tener "ideas" previas sobre qué es la inteligencia o, por lo menos, en que tareas se puede ver reflejada.

Cuando se les pregunta a la gente de la calle, la definición de inteligencia es distinta, ya que se hace más hincapié en las habilidades sociales. En un estudio, en el que se pedía a un grupo de no expertos que definiesen la inteligencia, se definía en términos de tres clases de competencias: (1) en la solución de problemas prácticos, (2) habilidad verbal, y (3) competencia social. Ahora bien, la definición de inteligencia varía según la ocupación del que la defina. Por ejemplo, un estudio encontró que para los profesores de filosofía lo más importante es el pensamiento crítico y lógico, mientras que los físicos daban más importancia al pensamiento matemático preciso, la habilidad de relacionar los fenómenos físicos con los conceptos físicos, y la habilidad de aprehender rápidamente las leyes de la naturaleza (Sternberg, 2003).

Por otra parte, la inteligencia es un concepto construido socialmente. Las culturas consideran "inteligentes" los atributos que favorecen el éxito dentro de esas culturas (Sternber y Kaufman, 1998). En las zonas rurales de Kenia, la inteligencia puede ser el don para discriminar cuáles son las hierbas naturales que sirven para curar enfermedades específicas. En las culturas asiáticas puede ser una habilidad social. En los países occidentales es el rendimiento superior en tareas cognitivas. En cada contexto la inteligencia es la capacidad de aprender a partir de la experiencia, de resolver problemas y de utilizar el conocimiento para adaptarse a situaciones nuevas.

En definitiva, al menos desde el punto de vista de los expertos existe un acuerdo general sobre el concepto de inteligencia. Pero, todavía quedan algunas controversias sin resolver: (1) ¿La inteligencia es una capacidad general única o varias capacidades específicas, (2) ¿Con las herramientas que ofrece actualmente la neurociencia podemos localizar la inteligencia en el cerebro y medirla?

EL PUNTO DE VISTA PSICOMÉTRICO O DEL ANÁLISIS FACTORIAL

Para descubrir si existe la posibilidad de que haya un factor general subyacente a nuestras capacidades mentales específicas, los psicólogos estudian el modo en que las distintas capacidades se relacionan entre sí. Un método estadístico llamado análisis factorial[3] permite que los investigadores identifiquen grupos de ítem del test que miden una capacidad común. Por ejemplo, las personas que tienen facilidad para los ítems de vocabulario suelen tener facilidad para la comprensión de historias, un grupo que ayuda a definir un factor de inteligencia verbal. Otros grupos incluyen un factor de capacidad espacial y un factor de capacidad de razonamiento.

En toda discusión sobre la inteligencia y sobre la medición de la inteligencia queda siempre en el aire la pregunta sobre la "naturaleza de la inteligencia". ¿Cómo se entiende esto que se mide y calcula de modo usual se relaciona con el éxito escolar o los grupos profesionales? ¿Hay una "inteligencia básica" que participa en todos los

[3] **Análisis factorial**: procedimiento estadístico multivariado que identifica grupos de elementos relacionados (denominados factores) en un test; se utiliza para identificar distintas dimensiones del resultado subyacentes a la puntuación global de una persona.

procesos intelectuales, o varias inteligencias que funcionan independientemente unas de otras?

La teoría más antigua sobre la organización y la estructura de la inteligencia provino de Charles Spearman (1904, 1923) que fue, a principios del siglo XX, uno de los primeros en investigar las mediciones de la inteligencia con la ayuda del análisis factorial. La idea surgió a partir de la observación de que determinadas personas respondían mejor a determinadas tareas (ítems) de los test de inteligencia que otras, a pesar de que todas ellas obtenían resultados iguales en el conjunto del test. Las tareas, por consiguiente, se diferencias entre sí en sus demandas, por lo que también miden capacidades distintas. A principios de este siglo, Spearman intentó demostrar con su modelo analítico por factores que para cada tipo de tarea del test de inteligencia era necesaria una capacidad intelectual general fundamental. Spearman denominó a esta capacidad intelectual general *factor de inteligencia general o factor g*. Este factor entra en el resultado de todas las mediciones de inteligencia. Junto con este factor general, un test de inteligencia mide otras *capacidades específicas del test*, también llamadas *factores s*, aunque son de poca importancia. Por consiguiente, el objetivo del desarrollo de test de inteligencia es encontrar tareas cuyos resultados obtengan una alta correlación con el patrón de resultados de una serie de otros test de inteligencia.

El psicólogo estadounidense L.L. Thurstone rechazaba las ideas de Spearman, ya que sostenía que la inteligencia se compone de siete capacidades mentales "primarias" (Thurstone, 1938): Habilidad Espacial (E), Memoria (M), Rapidez Perceptiva (P), Fluidez Verbal (F), Habilidad Numérica (N), Razonamiento (R) y Significado Verbal (V).

A diferencia de Spearman, pensaba que las habilidades anteriores son relativamente independientes. Así, una persona con una habilidad espacial excepcional podría carecer de fluidez verbal. Según Thurstone en conjunto las siete capacidades mentales primarias constituyen la inteligencia general.

La aparente contradicción entre los modelos de Sperman y Thurstone es más aparente que real, pues, técnicamente, en un análisis de segundo orden con una solución oblicua en lugar de ortogonal, aparecían factores que sugerían el factor "g" de Spearman (Sánchez-Cánovas, 1995).

En contraste con Thurstone, el psicólogo R.B. Catell (1971) identifica sólo dos grupos de capacidades mentales. El primero, que llama *inteligencia cristalizada*, abarca las habilidades como el razonamiento y las destrezas verbales y numéricas. Por ser las que se imparten en la escuela, Catell cree que la experiencia y la educación formal influyen profundamente en las puntuaciones obtenidas en las pruebas de inteligencia cristalizada. El segundo grupo lo integra lo que Catell llama *inteligencia fluida*, es decir, destrezas como la formación de imágenes espaciales y visuales, o la capacidad para percibir los detalles visuales y la memoria mecánica. La experiencia y la ecuación influyen menos en las puntuaciones obtenidas en las pruebas de este tipo de inteligencia.

La diferencia fundamental entre fluida y cristalizada radica en que los conceptos y las destrezas cognitivas adquiridas implicados en la primera reflejan experiencias relativamente comunes a todos los humanos, mientras que los conceptos y destrezas cognitivas adquiridas que definen inteligencia cristalizada representan de modo más inmediato el grado de inmersión en una cultura particular. De otra manera y según la propia definición de Catell: "inteligencia cristalizada satura sobre todo las destrezas

referidas a "juicios" adquiridos culturalmente, mientras que inteligencia fluida lo hace en ejecuciones en las que las diferencias individuales debidas a experiencias de aprendizaje desempeñan un papel muy escaso".

En consecuencia, inteligencia fluida tendrá un componente hereditario y biológico sustancial, mientras que inteligencia cristalizada se debería más a la historia del aprendizaje de cada individuo, pero ambas cooperando en cualquier tipo de ejecución y sometidas a determinantes comunes en grado diverso.

Por las mismas fechas, Guilford (1968) propuso un modelo claramente distinto a los anteriores. Sugiere que la inteligencia puede ser descrita mediante 150 factores distintos, que representan las diferentes combinaciones de las 5 operaciones (cognición, memoria, producción convergente, producción divergente y memoria), 5 contenidos (figurativo, simbólico, semántico y comportamental), y 5 productos (unidades, clases, relaciones, sistemas, trasformaciones e implicaciones).

De modo más preciso, dentro del parámetro **operación** la *cognición* hace referencia al descubrimiento, reconocimiento o comprensión. La operación de *memoria* introduce la información en el almacén de memoria con algún grado de permanencia. Debe distinguirse del almacén de memoria en sí mismo. Este último, según Guilford (1968) subyace a todas las operaciones; todas las aptitudes dependen de él. La *producción divergente* implica la generación de alternativas lógicas a partir de una información dada, dónde el énfasis se pone en la variedad, cantidad y relevancia del resultado a partir de la misma fuente. La *producción convergente* consiste en la generación de conclusiones lógicas a partir de una información dada. El énfasis se pone en la consecución del único o convencionalmente mejor resultado. La operación de *evaluación* implica la comparación de ítems de información determinando su bondad con respecto a los criterios lógicos adoptados, tales como identidad y consistencia.

La distinción entre información *figurativa* y *semántica*, dentro del parámetro **contenido**, es clara ya que la información figurativa tiene propiedades sensoriales, tales como la visual o la auditiva. Se le considera "concreta" en contraposición a la semántica de carácter "abstracto". La denominación de semántica se debe a que en este caso la información está vinculada con símbolos de palabras ordinariamente, aunque esta conexión no es esencial. La información *simbólica* alude a tests en los que usualmente se emplean letras o números. Son signos denotativos que no tienen significación en y por sí mismos. Guilford añadió la categoría *comportamental* basándose en la noción de Thorndike de "inteligencia social". Hace referencia a la clase de información que una persona deriva de observaciones de la conducta de otra persona. A partir de signos expresivos o del lenguaje corporal puede uno llegar a conocer los sentimientos, pensamientos o intenciones de otra persona. Es la información incluida en las interacciones humanas.

Por último, dentro del parámetro **producto**, distinguimos: *Unidades* o ítems de información relativamente circunscritos que tienen el carácter de "cosa". *Clases*: concepciones que subyacen a conjuntos de ítems de información agrupados en virtud de sus propiedades comunes. *Relaciones*: conexiones entre ítems de información basados en variables o puntos de contacto que se aplican a los mismos. *Sistemas*: agregados de ítems de informaciones organizados o estructurados; complejos de interrelaciones o partes que interactúan. *Transformaciones*: cambios de varias clases (redefiniciones, mutaciones, transiciones o modificaciones) en la información

existente. Incluyen cualquier tipo de cambio: movimiento en el espacio, reordenamiento o reagrupamiento de letras en palabras o simplificar una ecuación, redefinir una palabra o adaptar un objeto a un nuevo uso. La *implicación* es sugerida por otra información. Son conexiones circunstanciales entre ítems de información. La previsión o predicción depende de la capacidad de extrapolar desde una información dada alguna condición o evento que naturalmente se sigue. La expresión "si…entonces" describe una implicación (ver Sánchez-Cánovas, 1984).

Como señala Sternberg (2003), no todos los trabajos dentro de la orientación psicométrica o factorial tienen una clara base teórica. Una de las primeras teorías de la inteligencia fue propuesta por el inglés Sir Francis Galton (1883). Galton creía que la inteligencia es la capacidad para trabajar y sensibilizarse con los estímulos externos. Galton y su seguidor en EEUU, James M. Catell, midieron la inteligencia usando tests que evalúan habilidades manuales, perceptivas o de amplitud de memoria.

Galton conocía muy bien los trabajos realizados en el campo de la psicofisiología sensorial; y muy en concreto los referidos a la medida de los tiempos de reacción. De hecho, vino a plantear que la medida de los tiempos de reacción podía ser un adecuado procedimiento de medida de la aptitud natural (inteligencia). En su famoso laboratorio antropométrico introdujo una serie de instrumentos: el silbato de Galton, que determinaba el grado de sensibilidad auditiva para los tonos altos; una barra para establecer el alcance visual; un aparato para medir los tiempos de reacción y, por último, la mesa de desayuno para medir la capacidad manipulativa. Si bien inicialmente concentró su interés en la medida de las características físicas y fisológicas, pronto pretendió hacer lo mismo con las diferencias psicológicas, eligiendo el experimento de tiempo de reacción como técnica básica (Tortosa, 1998).

TEST DE INTELIGENCIA

A pesar de la controversia en la que se han visto envuelto en muchas ocasiones, no cabe duda de que el uso de los test está unido indisolublemente a los inicios del estudio de la inteligencia desde la psicología. El primer test de inteligencia lo diseñaron Alfred Binet y su colaborador, Theodor Simon, para el sistema escolar público francés a principios del siglo XX (1905). Binet y Simon, desde el laboratorio de psicología de la Sorbona, desarrollaron una serie de preguntas y las probaron con los alumnos de París, con el objetivo de localizar a los niños retardados o con problemas de aprendizaje.

La primera escala de Binet-Simon se publicó en 1905. Consistía en 30 test ordenados por grado de dificultad. El examinador empezaba desde el inicio de la lista con cada niño, prosiguiendo poco a poco hasta que el niño ya no respondía a las preguntas. Para 1908 habían evaluado suficientes niños como para predecir el promedio que debía tener un niño de acuerdo con cada nivel de edad. A partir de estas puntuaciones Binet desarrolló el concepto de edad mental. Por ejemplo, un niño que obtiene una puntuación igual al promedio correspondiente a los años tiene una edad mental de 4.

En los siguientes 10 años se publicaron numerosas adaptaciones de Binet. La más conocida fue preparada en la Universidad de Stanford por L.M. Terman y se publicó en 1916. Terman fue el que introdujo el término de **cociente intelectual** (CI), con el fin de establecer un valor numérico para la inteligencia (ver figura 3). La escala de

Stanford-Binet se ha revisado en diversas ocasiones desde 1916. Los distintos subtests de la escala están diseñados para medir las cuatro clases de habilidades mentales que se consideran, casi universalmente, características de la inteligencia: razonamiento verbal, razonamiento abstracto/visual, razonamiento cuantitativo y memoria a corto plazo.

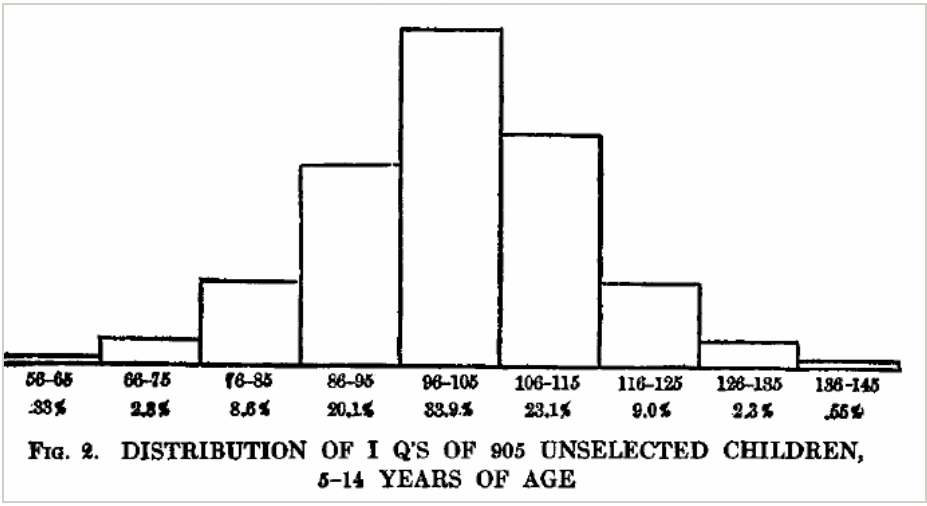


FIG. 2. DISTRIBUTION OF I Q'S OF 905 UNSELECTED CHILDREN, 5–14 YEARS OF AGE

Figura 3: Representación gráfica de la distribución del CI, en la obra de Terman (1916)

Sin embargo, el test individual para adultos que más se utiliza en la actualidad es la **Escala de Inteligencia de Weschler para Adultos** (Weschler Adult Intelligence Scale-Revised, WAIS-R). El WAIS original lo desarrolló David Weschler, un psicólogo del hospital Bellevue de Nueva York. Weschler puso objeciones al uso de la escala de Stanford-Binet en adultos en tres aspectos fundamentales. Primero, los problemas fueron diseñados para niños y les parecerían infantiles a los adultos. Segundo, las normas de edad mental de la escala Stanford-Binet no so aplicables a los adultos. Por último, la escala de Stanford-Binet pone el énfasis en las habilidades verbales y Weschler creía que la inteligencia adulta consiste más en la capacidad para manejar el medio ambiente que en la capacidad para solucionar problemas abstractos y verbales.

El WAIS-R se divide en dos partes, una de habilidades verbales y otra manipulativas o de ejecución. Aunque el contenido del WAIS-R es un poco más elaborado que el de la escala Stanford-Binet, la principal aportación de Weschler es su sistema de puntuación. Primero, a la persona se le otorgan calificaciones independientes de la parte verbal y de la manipulativa, así como una calificación del CI global. Segundo, en algunas pruebas la persona puede obtener uno o dos puntos extra, dependiendo de la complejidad de la respuesta dada. Este original sistema de calificaciones otorga crédito a las cualidades reflexivas que esperaríamos encontrar en los adultos inteligentes. Tercero, en algunas pruebas tanto la velocidad como la exactitud afectan a la puntuación que se obtiene.

Weschler también desarrolló un test de inteligencia similar al anterior para utilizarlo con niños en edad escolar. Al igual que el WAIS-R, la versión de 1991 de la **Escala de Inteligencia de Weschler para Niños** (Weschler Intelligence Scale for Children-Third Edition, WISC-III) proporciona calificaciones independientes para la parte verbal y la de ejecución, además de una calificación de CI global.

La WAIS-R consta de once subtests, los seis primeros del CI verbal y los cinco restantes del CI manipulativo o "de ejecución".

Información general. Contiene ítems del tipo de "¿Quién es el Presidente de los Estados Unidos, pero se trata en este caso de un ítem fácil, de "introducción", que en realidad no se puntúa.

Comprensión. Los ítems preguntan típicamente: "¿Qué harías si…?" "¿Por qué hacemos corrientemente…?", rozando el "sentido común" y el conocimiento de las costumbres sociales. Estos ítems dan lugar a más dificultades que los ítems del primer subtest. Se dan instrucciones detalladas al examinador.

Razonamiento aritmético. Problemas sencillos de aritmética mental.

Dígitos. Se leen en voz alta series de dígitos (números de una sola cifra). En la primera parte, se pide al sujeto que repita las series; en la segunda que las repita a la inversa.

Semejanzas. Se leen en voz alta pares de palabras. En cada caso, el sujeto tiene que decir en qué forma se parecen las dos cosas.

Vocabulario. Al sujeto se le pide que explique el significado de las palabras, presentadas en orden creciente de dificultad. Se facilitan ejemplos para ayudar al examinador a que puntúe definiciones dudosas.

Clave de números. Se adaptó en principio de la Escala de Ejecución del Ejército, usada por el ejército de los EEUU en la primera guerra mundial. Los símbolos y los números se presentan emparejados y el sujeto tiene que continuar emparejando símbolos con los correspondientes números.

Figuras incompletas. Se presentan imágenes en las que falta una parte (por ejemplo, la nariz de una cara). Se pide al sujeto que nombre la parte que falta.

Cubos. Similar al Diseño de Bloques de Kohs.

Historietas. Se presentan juegos de tarjetas, uno cada vez, conteniendo cada uno imágenes que por su correcto orden constituyen una historia. El sujeto tienen que colocar cada juego por su debido orden.

Rompecabezas. El sujeto tiene que formar objetos (por ejemplo, un ser humano) colocando las piezas en la debida posición a la manera de un rompecabezas.

Las Matrices Progresivas de Raven, fueron creadas por Raven (1938), y fue un test pensado para evaluar a un grupo selectivo de personas (los oficiales de la armada estadounidense). Basada en el antecedente de Raven y Penrose (1936), esta prueba tiene que ver con el razonamiento analógico, la percepción y la capacidad de abstracción. Se trata de un test no verbal, donde el sujeto tiene que encontrar la figura que falta en una serie de láminas pre-impresas. Se pretende que el sujeto utilice habilidades perceptuales, de observación y razonamiento analógico para deducir el elemento que completa la matriz gráfica. A la persona se le pide que analice la serie que se le presenta y que siguiendo la secuencia horizontal y vertical, escoja uno de los ocho trazos: el que encaje perfectamente en ambos sentidos, tanto en el horizontal como en el vertical. Casi nunca se utiliza límite de tiempo, pero dura aproximadamente 60 minutos.

De entre la gran cantidad de test que existen para evaluar la inteligencia, y por supuesto sus diferentes componentes, capacidades o habilidades, es necesario destacar el test de Raven, ya que es un instrumento para medir la capacidad intelectual independientemente de los conocimientos adquiridos. De esta manera brinda información sobre la capacidad y claridad de pensamiento presente del examinado para la actividad intelectual, en un tiempo ilimitado.

Existen tres versiones diferentes de la prueba, la más usual es la Escala General (12 elementos en 5 series A, B, C, D, E), para sujetos de 12 a 65 años, donde la complejidad aumenta de manera progresiva. También están las Matrices Progresivas en Color, incluyendo una versión para niños, y las Matrices Superiores, para personas con mayor capacidad.

La puntuación del Test de Matrices Progresivas de Raven (escala general), independientemente del tiempo ocupado en realizarlo, evalúa la capacidad intelectual general del individuo, lo que tendría una relación directa con la velocidad con que un sujeto procesa la información y con su capacidad de atención. En un trabajo realizado en esta línea, la correlación inversa encontrada entre la puntuación del test de Raven y la latencia de la onda P300 significa que la velocidad con la que los sujetos reconocieron el estímulo *diana* (target) se correlaciona con su capacidad intelectual, es decir, a mayor puntuación en el test, mayor velocidad de procesamiento mental (De Bortoli, Barrios y Azpiroz, 2002).

Quizá la cuestión más controvertida del uso de los test de inteligencia es la carga lingüística y cultural que pueden tener. Una vez superadas las controvertidas polémicas sobre la raza y la inteligencia o sobre la influencia de la herencia y el medio, en la actualidad se entiende que cuando se mide la inteligencia hay que tener ambas cargas en cuenta: la cultural y la lingüística. Como se puede ver en la Figura 4, ambas demandas interactúan entre sí y sólo cuando ambas son bajas, podemos hablar de que la ejecución (en un determinado test) se ve menos afectada. En los test que hemos utilizado como ejemplo en este trabajo, WAIS y Raven, podemos encontrar esta cuestión. En el WAIS, los test que componen la escala verbal, tendrían una importante demanda lingüística y cultural; sin embargo, las demandas de los test de la escala manipulativa son mucho menores. El grado de demanda tanto cultural, como sobretodo lingüística del test de matrices de Raven es bajo (Ortiz y Ochoa, 2005).

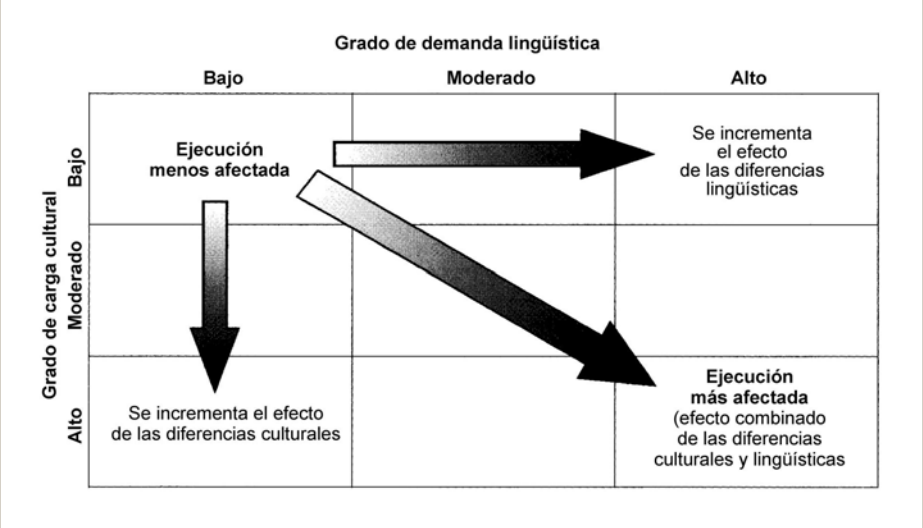


Figura 4: Influencia de la interacción entre el grado de demanda lingüística y cultural en la ejecución de un test de inteligencia (adaptado de Ortiz y Ochoa, 2005).

LA INTELIGENCIA DESDE LA PSICOLOGÍA COGNITIVA: EVIDENCIA DE LA ORGANIZACIÓN ESTRUCTURAL

Una vez que hemos llegado hasta aquí, parece quedar claro que el término inteligencia agrupa a un conjunto amplio de capacidades mentales más o menos relacionadas entre sí. Los resultados acumulados de 100 años de investigación en las covariaciones de los test, tareas y paradigmas diseñados para identificar los rasgos fundamentales de la inteligencia humana, indican que existen no menos de 87 capacidades elementales diferentes (Horn y Blakson, 2005).

Utilizando la misma evidencia, basada en el análisis factorial, que se ha utilizado para hallar los rasgos de primer orden, se puede decir que existen ocho (o nueve) factores de segundo orden que reúnen a los factores de primer orden.

Mucho de lo que sabemos acerca del desarrollo de las habilidades, y la mayoría de las teorías acerca de la naturaleza de la inteligencia humana tienen que ver con las habilidades de segundo orden. Estas pueden describirse brevemente de la siguiente forma[4]:

Conocimiento Aculturación *(Gc)*[5], medido por test que indican la amplitud y profundidad del conocimiento del lenguaje, conceptos e información de la cultura dominante.

[4] Para mayor detalle sobre las habilidades de primer orden que se incluyen en cada uno de los ocho factores, ver Horn y Blakson (2005), páginas 44 y 45.

[5] En este apartado se introduce la nomenclatura en inglés que se utiliza como denominador de las distintas habilidades cognitivas tanto de primer o segundo orden como en el estrato I de Carroll (2005).

Razonamiento Fluido *(Gf)*, medido por tareas que requieren razonamiento. Indican la capacidad para identificar relaciones, aprehender implicaciones y hacer inferencias tanto en un contexto nuevo como en uno familiar.

Aprehensión y recuperación a corto plazo *(SAR)*, también denominado *memoria a corto plazo (Gsm)* y *memoria de trabajo*. Se mide mediante una variedad de tareas que requieren mantener conocimiento los elementos de una situación inmediata (p.e. el span un minuto o tanto).

Fluidez de la recuperación de lo almacenado a largo plazo *(TSR)*, también llamada *memoria a largo plazo (Glm)*. Se mide mediante tareas que indican consolidación del almacenamiento y tareas que requieren la recuperación mediante asociaciones de información almacenadas minutos, horas, semanas o años después.

Velocidad de procesamiento *(Gv)*, implicada en casi todas las tareas intelectuales y se puede medir en tareas simples de detección y comparación rápida en las que si no fuese por la alta velocidad la mayoría de las personas detectaría la respuesta correcta.

Procesamiento Visual *(Gv)*, medido en tareas que requieren cierre visual, constancia y fluidez a la hora de reconocer la forma de los objetos que aparecen en el espacio visual, que han sido previamente rotados o cambiados.

Procesamiento Auditivo *(Ga)*, medido por tareas que requieren la percepción de patrones de sonidos bajo condiciones de distracción o distorsión, el mantenimiento de la conciencia del orden y el ritmo entre los sonidos y la captación de los elementos de grupos de sonido.

Conocimiento Cuantitativo *(Gq)*, medido en tareas que requieren la comprensión y la aplicación de los conceptos y habilidades matemáticas.

Una estructura prácticamente idéntica es la propuesta por Carroll (2005) a la que denomina *Teoría de las Habilidades Cognitivas de los Tres Estratos* (Three Stratum) y que es una expansión y extensión de las teorías previas. Especifica los distintos tipos de habilidades cognitivas que existen y cómo están relacionadas unas con. Proporciona un mapa de las habilidades cognitivas, como se puede ver con detalle en la figura 5. De la inteligencia general (**Estrato III, General**) se derivan ocho conjuntos extensos de habilidades (**Estrato II, Extenso**) y de éstos, a su vez se derivan todas y cada una de las habilidades cognitivas limitadas o primarias (**Estrato I, Limitado**).

Los ocho componentes del estrato II (extenso) serían prácticamente idénticos a los que acabamos de exponer: Inteligencia Fluida, Inteligencia Cristalizada, Memoria General y Aprendizaje, Percepción Auditiva Extensa, Habilidad de Recuperación Extensa, Velocidad Cognitiva Extensa y Velocidad de Procesamiento (TR, Velocidad de Decisión).

La **Inteligencia Fluida** se concretaría en Razonamiento Secuencial General (RG), Inducción (I) o Razonamiento Cuantitativo (RE).

La **Inteligencia Cristalizada** incluiría Comprensión Lectora (RC), Habilidad de Completar (CZ), Decodificación Lectora (RD), Velocidad Lectora (RS), Habilidad Ortográfica (SG), Habilidad de Escritura (WA), Aptitud para Lenguas Extranjeras (LA), Desarrollo del Lenguaje (LD), Conocimiento del Léxico (VL), Habilidad de Escucha (LS), Codificación Fonética (PC), Habilidad de Comunicación (CM), Producción Oral y Fluidez (OP), Sensibilidad Gramatical (MY) y Comprensión del Lenguaje Verbal (V).

La **Memoria General y Aprendizaje** en el estrato I se desglosa en Espacio de Memoria (MS), Memoria Asociativa (MA), Memoria con Significado (MM), Memoria de Recuerdo Libre (M6), Memoria Visual (MV) y Habilidades de Aprendizaje (L1).

La **Percepción Visual Extensa** tiene que ver con Integración Perceptiva Serial (PI), Estimación del Longitud (LE), Percepción de Ilusiones Visuales (IL), Percepción de Alternaciones (PN), Imaginación (IM), Visualización (VZ), Relaciones Espaciales (SR), Velocidad de Cierre (CS), Flexibilidad de Cierre (CF), Velocidad Perceptiva (P) y Escaneo Espacial (SS).

La **Percepción Auditiva Extensa** se ve reflejada en Mantener y Juzgar el Ritmo (U8), Discriminación de la Intensidad/Duración del Sonido (U6), Discriminación de la Frecuencia del Sonido (U5), Umbral Auditivo (UA, UT, UU), Tono Absoluto (UP), Localización de Sonidos (UL), Discriminación de Sonidos del Habla (US), Discriminación de Sonidos General (U3), Resistencia a la Distorsión de Estímulos Auditivos (UR), Rastreo Temporal (UK), Memoria para Patrones de Sonido (UM) y Discriminación y Juicio Musical (U1, U9).

La **Habilidad de Recuperación Extensa** se subdivide en Originalidad/Creatividad (FO), Sensibilidad a los Problemas (SP), Fluidez Figurativa (FF), Flexibilidad Figurativa (FX), Fluidez de Ideas (FI), Fluidez Asociativa (FA), Fluidez Expresiva (FE), Facilidad Denominativa (NA), Fluidez de Palabras (FW).

La **Velocidad Cognitiva Extensa** se relaciona con la Rapidez en Captar los Test (R9), la Facilidad Numérica (N) y la Velocidad Perceptiva (P).

Por último, la **Velocidad de Procesamiento** (Tiempo de Reacción, Velocidad de Decisión) se puede medir mediante el Tiempo de Reacción Simple (R1), el Tiempo de Reacción en la Elección (R2), la Velocidad de Procesamiento Semántico (R4) y la Velocidad de Comparación Mental (R7).

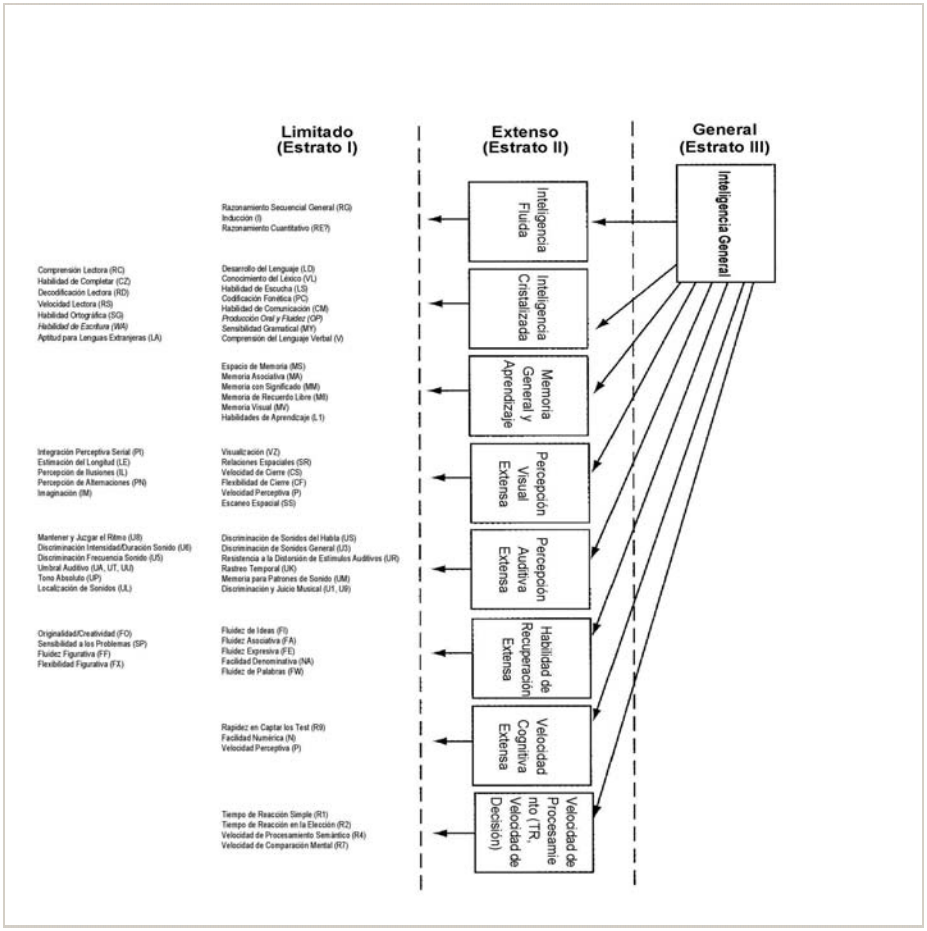


Figura 5: La estructura de los tres estratos de las habilidades cognitivas (adaptado de Carroll, 2005).

LA INTELIGENCIA EMOCIONAL

Algo diferente a la inteligencia 'académica' es lo que Cantor y Kihlstrom (1987) denominaron *inteligencia social*, el conocimiento necesario para comprender las situaciones sociales y manejarse a sí mismo con éxito. En ese sentido, en diversas ocasiones, se ha mostrado que el rendimiento académico en la universidad tiene muy poca capacidad predictiva sobre el éxito laboral posterior (McClelland, 1993).

Quizá la pregunta que debemos hacernos es porqué las personas con una capacidad intelectual superior no son mejores a la hora de desenvolverse mejor en la vida y obtener un mejor bienestar físico y mental. La respuesta puede estar en lo que Salovey y Mayer denominan **inteligencia emocional** (Salovey y Mayer, 1990): la capacidad para percibir, expresar, entender y regular las emociones. La inteligencia emocional es una parte de la inteligencia social y no tiene que estar necesariamente con la aptitud académica. La IE fue formalmente definida en 1990 como "*una parte de la*

inteligencia social que incluye la capacidad de controlar nuestras emociones y la de los demás, discriminar entre ellas y usar dicha información para guiar nuestro pensamiento y nuestros comportamientos" (Salovey y Mayer, 1990)

Las personas emocionalmente inteligentes son conscientes de sí mismas. Pueden manejar sus emociones sin verse invadidos por una depresión, ansiedad o enfado incontrolables. Pueden retrasar una gratificación en la búsqueda de grandes recompensas, antes que verse llevadas por sus impulsos. Su empatía les permite leer las emociones de los demás. Pueden manejar las emociones ajenas con habilidad y controlar de este modo los conflictos. En pocas palabras, son listos emocionalmente y por ello a menudo éxito en sus carreras profesionales, en sus matrimonios y en su condición de padres, ámbitos en los que fracasan otras personas más inteligentes desde el punto de vista académico (pero menos inteligentes desde el punto de vista emocional).

Mayer, Salovey y Caruso (2000) desarrollaron recientemente la Escala Multifactorial de Inteligencia Emocional (MEIS) para evaluar la inteligencia emocional general y sus tres componentes. Los ítems evalúan la habilidad del entrevistado para:

Percibir las emociones al reconocer las emociones trasmitidas por los rostros, fragmentos musicales, diseños gráficos y cuentos.

Entender las emociones al reconocer cómo cambian a través del tiempo, predecir emociones diferentes (por ejemplo, lo que siente un conductor que atropella a un perro que corre tras un palo y lo que siente el dueño del perro), y aprehender cómo se combinan. (Por ejemplo: "¿Cuáles son las dos emociones que más se unen al optimismo? El placer y la anticipación, la aceptación y la alegría, la sorpresa y la alegría, el placer y la alegría").

Regular las emociones al clasificar las estrategias alternativas que una persona puede utilizar cuando se enfrenta a conflictos de la vida real.

La importancia de los conceptos anteriores, compresión y regulación de los estados emocionales, ha quedado manifiesta en la teoría de la Inteligencia Emocional (IE) desarrollada por Salovey y colaboradores (Mayer y Salovey, 1993; Mayer y Salovey, 1997; Salovey y Mayer, 1990; Salovey, Hsee y Mayer, 1993). La IE se entiende como la habilidad de los sujetos para procesar y utilizar la información proporcionada por las emociones con fines adaptativos. A partir del concepto de Inteligencia Personal de Gardner (1983), entendida en parte como la capacidad de discriminar inmediatamente los propios sentimientos, etiquetarlos, simbolizarlos y utilizarla como un medio para comprender y guiar nuestro comportamiento, Salovey y colaboradores han definido la IE como la capacidad de identificar nuestros sentimientos y los de los demás, regularlos y utilizar esa información para conseguir una conducta más adaptativa. Un importante aspecto de la IE es la habilidad para prestar atención a los propios sentimientos y los de los demás, la claridad con la que se experimentan esos sentimientos y las creencias acerca de cómo terminar los estados de ánimo negativos o prolongar los positivos. A estos procesos de inteligencia emocional percibida (IEP) se les ha llamado *meta-mood experience* y se ha desarrollado para su medición la escala TMMS (*Trait Meta-Mood Scale*, Salovey et al., 1995) que incluye tres componentes: Atención que se presta a los propios sentimientos y a los de los demás; Claridad, definida como la capacidad para discernir los propios sentimientos; y

Regulación, que es la capacidad para cambiar una emoción negativa o mantener una positiva (Salovey, Mayer, Goldman, Turvey y Palfai, 1995).

Actualmente, se puede decir que existe bastante evidencia acumulada a favor de la IE como un modelo explicativo, para estudiar las diferencias individuales relacionadas con el procesamiento de la información emocional (Salovey y Grewal, 2005). A pesar de ello, como señalan Extremera, Fernández-Berrocal, Mestre y Guil (2005), el debate sobre la inteligencia emocional humana está aún en sus inicios. Son numerosos los científicos que muestran su desconfianza hacia el hecho de que la IE pueda ser evaluada por medio de test de papel y lápiz. Otros tantos ponen en duda la propia naturaleza empírica del constructo y sus detractores resaltan su parecido con otras inteligencias similares (p.e. inteligencia exitosa, inteligencia social, inteligencia práctica).

OPCIONES RECIENTES EN LA CONCEPCIÓN DE LA INTELIGENCIA

Howard **Gardner** (1983, 1993) coincide con Thurstone en que la inteligencia viene en paquetes distintos. Observa que la lesión cerebral puede mermar algún tipo de capacidad pero no otras. Gardner también estudia los informes de personas con capacidades escepcionales, entre ellas las que destacan en una única capacidad. Las personas con el *sindrome del sabio*[6], por ejemplo, obtienen puntuaciones mínimas en los tests de inteligencia pero poseen alguna capacidad increíble en la informática, el dibujo o la memoria musical, entre otros. Es posible que dichas personas prácticamente no tengan capacidad verbal, pero que sean capaces de calcular números con la misma rapidez y precisión que una calculadora electrónica o de identificar casi de forma instantánea el día de la semana que corresponde a cualquier fecha de la Historia.

A partir de estos datos, Gardner defiende que no tenemos *una* inteligencia, sino que en su lugar poseemos *múltiples* inteligencias, independientes las unas de las otras.

La teoría de las inteligencias múltiples (IM) pluraliza el concepto tradicional de inteligencia. Una inteligencia implica la habilidad necesaria para resolver problemas o para elaborar productos que son de importancia en un contexto cultural o en una comunidad determinada. La capacidad para resolver problemas permite abordar una situación en la cual se persigue un objetivo, así como determinar el camino adecuado que conduce a dicho objetivo (Gardner, 1993).

La teoría de las IM se organiza a la luz de los orígenes biológicos de cada capacidad para resolver problemas. Sólo se tratan las capacidades que son universales a la especie humana. Aún así, la tendencia biológica a participar de una forma concreta de resolver problemas tiene que asociarse también al entorno cultural. Puesto que deseamos seleccionar inteligencias que estén enraizadas en la biología, que sean valoradas en uno o varios contextos culturales, ¿cómo se identifica realmente una "inteligencia"? En su obra *Frames of Mind* (1983) responde a esta pregunta, ya que contiene una discusión de los criterios que debe cumplir una "inteligencia".

Los criterios incluyen evidencias procedentes de fuentes distintas: a) conocimiento acerca del desarrollo normal y del desarrollo en individuos superdotados; b)

[6] **Síndrome del sabio**: condición por la que una persona con una capacidad mental limitada posee una sorprendente capacidad específica.

información acerca del deterioro de las capacidades cognitivas a consecuencia de una lesión cerebral; c) estudio de poblaciones excepcionales; d) datos de la evolución de la cognición a través de la historia de la humanidad; e) estimación de la cognición a través de las culturas; f) estudios psicométricos; y g) estudios psicológicos de aprendizaje, particularmente las medidas de transferencia y generalización entre tareas (Gardner, 1993).

Además de cumplir los criterios anteriores, cada inteligencia debe poseer una operación nuclear identificable, o un conjunto de operaciones. Como sistema computacional basado en las neuronas, cada inteligencia se activa o se dispara a partir de ciertos tipos de información interna o externa. Por ejemplo, el núcleo de la inteligencia musical es la sensibilidad para entonar bien, mientras que un núcleo de la inteligencia lingüística es la sensibilidad para los rasgos fonológicos.

Una inteligencia debe ser también susceptible de codificarse en un sistema simbólico. De hecho, la existencia de una capacidad computacional nuclear anticipa la existencia de un sistema simbólico que aproveche esta capacidad. Aunque es posible que una inteligencia funcione sin un sistema simbólico, su tendencia a una formalización de este tipo constituye una de sus características primarias.

Gardner propone la existencia de siete inteligencias distintas: inteligencia lógico-matemática, inteligencia lingüística, inteligencia espacial, inteligencia musical, inteligencia cinestésico-corporal, inteligencia interpersonal e inteligencia intrapersonal. Las dos primeras son conocidas porque forman parte de las otras teorías de la inteligencia que hemos visto.

Así, la **Inteligencia Lógico-Matemática** implica la capacidad para emplear los números de manera efectiva y de razonar adecuadamente a través del pensamiento lógico. Comúnmente se manifiesta cuando trabajamos con conceptos abstractos o argumentaciones de carácter complejo. Cuando se enfrentan a problemas complejos, las personas que tienen un nivel alto en este tipo de inteligencia poseen sensibilidad para realizar esquemas, relaciones lógicas, afirmaciones, proposiciones, funciones y otras abstracciones relacionadas. Un ejemplo de ejercicio intelectual de carácter afín a esta inteligencia es resolver test de cociente intelectual.

La **Inteligencia Lingüística** describe la capacidad para el lenguaje hablado y escrito, la habilidad para aprender idiomas, comunicar ideas y, lograr metas usando la capacidad lingüística. Esta inteligencia incluye también la habilidad de usar efectivamente el lenguaje para expresarse retóricamente o tal vez poéticamente; esta inteligencia es frecuente en escritores, poetas, abogados, líderes carismáticos y otras profesiones que utilizan sobre otras habilidades la de comunicarse efectivamente.

La **Inteligencia Espacial** es la capacidad de pensar en tres dimensiones. Permite percibir imágenes externas e internas, recrearlas, transformarlas o modificarlas, recorrer el espacio o hacer que los objetos lo recorran y producir o decodificar información gráfica. Presente en pilotos, marinos, escultores, pintores y arquitectos, entre otros. Está en los alumnos que estudian mejor con gráficos, esquemas, cuadros. Les gusta hacer mapas conceptuales y mentales. Entienden muy bien planos y croquis.

La **Inteligencia Musical** es la capacidad de percibir, discriminar, transformar y expresar las formas musicales. Incluye la sensibilidad al ritmo, al tono y al timbre. Está presente en compositores, directores de orquesta, críticos musicales, músicos, *luthiers* y oyentes sensibles, entre otros. Los alumnos que la evidencian se sienten

atraídos por los sonidos de la naturaleza y por todo tipo de melodías. Disfrutan siguiendo el compás con el pie, golpeando o sacudiendo algún objeto rítmicamente.

La **Inteligencia Cinestésico-Corporal** es la capacidad para usar todo el cuerpo en la expresión de ideas y sentimientos, y la facilidad en el uso de las manos para transformar elementos. Incluye habilidades de coordinación, destreza, equilibrio, flexibilidad, fuerza y velocidad, como así también la capacidad cinestésica y la percepción de medidas y volúmenes. Se manifiesta en atletas, bailarines, cirujanos y artesanos, entre otros. Se puede apreciar en los alumnos que destacan en actividades deportivas, danza, expresión corporal y/o en trabajos de construcciones en los que se utilizan distintos materiales. También en aquellos que son hábiles en la ejecución de instrumentos.

La **Inteligencia Interpersonal** es la capacidad de entender a los demás e interactuar eficazmente con ellos. Incluye la sensibilidad a expresiones faciales, la voz, los gestos y posturas y la habilidad para responder. Presente en actores, políticos, buenos vendedores y docentes exitosos, entre otros. La tienen los alumnos que disfrutan trabajando en grupo, que son convincentes en sus negociaciones con compañeros y profesores, que entienden al compañero.

La **Inteligencia Intrapersonal** es la capacidad de construir una percepción precisa respecto de sí mismo y de organizar y dirigir su propia vida. Incluye la autodisciplina, la autocomprensión y la autoestima. Se encuentra muy desarrollada en teólogos, filósofos y psicólogos, entre otros. La evidencian los alumnos que son reflexivos, de razonamiento acertado y suelen ser consejeros de sus compañeros y amigos..

El punto más fuerte de Gardner es que nuestras aptitudes mentales incluyen más que las que nos permiten aprobar en la universidad. Muchas teorías sobre la inteligencia sitúan a las capacidades lingüística y matemática en un pedestal. Por el contrario, Gardner presta atención a una gama más amplia de capacidades.

En la misma línea, Robert Sternberg (1984, 1985, 1996, 2005) ha propuesto una **teoria triárquica de la inteligencia**, según la cual la inteligencia humana comprende una variedad mucho más amplia de habilidades que las imaginadas por los teóricos anteriores y que las habilidades necesarias para un desempeño eficaz en el mundo real son tan importantes como las habilidades más limitadas evaluadas por los tests de inteligencia tradicionales. En este sentido, la teoría de la inteligencia de Sternberg es muy cercana al punto de vista informal que el común de las personas sostiene sobre la inteligencia.

Sternberg (1985, 1996) coincide con la idea de Gardner de las inteligencias múltiples, pero distingue simplemente tres tipos de inteligencia:

Inteligencia analítica (resolución de problemas académicos), evaluada por los test de inteligencia, que presentan problemas bien definidos con una respuesta correcta única.

Inteligencia creativa, demostrada en la reacción adaptativa frente a situaciones nuevas y la producción de nuevas ideas.

Inteligencia práctica, necesaria para las tareas cotidianas, que suelen estar mal definidas y presentan muchas soluciones.

Los tradicionales test de inteligencia evalúan la inteligencia académica. Predicen los resultados escolares con bastante exactitud pero no resultan fiables en cuanto a la predicción del éxito profesional. Las personas que muestran una alta inteligencia práctica pueden no haber destacado en la escuela. El ser un buen director no depende,

por ejemplo, tanto de las capacidades académicas evaluadas mediante la puntuación obtenida en un test de inteligencia (suponiendo que esté en torno a la media o por encima de ella) como de la habilidad para gestionarse a sí mismo, a las propias tareas y a tras personas.

El test de Sternberg y Wagner (1993, 1995) sobre la inteligencia de gestión práctica mide la capacidad de la persona que realiza el test para redactar buenos informes, motivar a los otros, saber cuándo delegar tareas y responsabilidades, comprender a las personas y cómo promocionarse a sí mismos. Los ejecutivos de empresas que obtienen puntuaciones más altas tienden a ganar salarios superiores y obtienen mejores rendimientos que los que obtienen puntajes bajos.

Aunque Sternberg (1998, 1999) y Gardner (1998) no coinciden en ciertos puntos específicos, están de acuerdo en que múltiples habilidades pueden contribuir al éxito en la vida. También aceptan que las diferentes variedades de talentos añaden diversidad y desafíos a la educación. Bajo la influencia de Gardner o Sternberg, muchos maestros se han entrenado para valorar las diferencias en la capacidad y aplicar la teoría de la inteligencia múltiple en sus clases. Las evaluaciones de estos programas están en marcha (Myers, 2004).

PSICOBIOLOGÍA DE LA INTELIGENCIA

Aunque se muestren correlaciones entre la anatomía cerebral y la inteligencia, es difícil que estas correlaciones puedan explicar las diferencias en inteligencia. Sin embargo, el rapidísimo avance de las neurociencias ha aportado mucha luz a los conocimientos en el área de la psicobiología de la inteligencia.

Quizá el modelo que mejor integra los conocimientos de la neurociencia sobre las bases neurales de la inteligencia es el expresado en el trabajo de Newman y Just (2005). La propuesta principal del modelo, para explicar la inteligencia fluida (g_f), es que lo bien que el sistema neural se adapte a los cambios en el medio ambiente puede afectar a la calidad y la eficiencia del procesamiento, constituyéndose, de ese modo, en la principal fuente de diferencias individuales. A modo de resumen, la teoría se compone de cuatro principios de la computación cortical:

Capacidad de procesamiento. Durante la ejecución de tareas cognitivas se consume energía, teniendo cada área cortical una capacidad de recursos limitada. Este principio tiene implicaciones directas para las diferencias individuales en inteligencia. Primero, sugiere que la cantidad de recursos disponibles dentro del sistema neural varía de unos individuos a otros. Segundo, la cantidad de recursos que se requieren para ejecutar una tarea pueden variar entre los sujetos debido a variaciones en la eficiencia.

Maleabilidad de las redes de procesamiento. La topología (composición cortical) de las redes neurocognitivas asociadas a una tarea dada cambia dinámicamente, adaptándose ella misma a las demandas de la tarea dada. Por tanto, la eficiencia con la que este cambio topológico ocurre puede contribuir a las diferencias individuales en la ejecución de la tarea.

Conectividad funcional. Las regiones corticales funcionan colaborando unas con otras en la ejecución de tareas. La variación en el grado de sincronización y eficiencia en la comunicación entre regiones puede contribuir a las diferencias individuales en la ejecución de tareas.

Conectividad anatómica. La calidad de las vías de materia blanca que conectan las áreas corticales puede influir también en la velocidad de procesamiento. La variación en el grado o calidad de las conexiones anatómicas entre las regiones de procesamiento pueden contribuir a las diferencias individuales en la ejecución de tareas.

Estos cuatro principios se asientan en diferentes aportaciones empíricas basadas en la neuroimagen funcional, para mayor detalle ver el trabajo de Newman y Just (2005). En cualquier caso, las aportaciones de las neurociencias a la psicología de la inteligencia se circunscriben fundamentalmente al estudio de la Inteligencia Fluida. Sin embargo, otras áreas más cercanas a la psicología experimental también están contribuyendo al desarrollo de la investigación sobre las bases neurales de la inteligencia como es el caso de la denominada cronometría mental.

La cronometría mental puede definirse como la medición de la velocidad cognitiva. Esto es, el tiempo que se tarda en procesar la información de diferentes tipos y grados de complejidad. Las medidas básicas son los tiempos de reacción (TR) a estímulos visuales o auditivos que llaman a una respuesta particular, elección o decisión.

Desde los trabajos de Galton (1822-1911), el padre de la psicología diferencial, diversos autores han venido planteando que la velocidad mental es el aspecto central de la inteligencia general. En ese sentido, los defensores de la cronometría mental estiman que ésta supone un importante avance a la hora de acercar la psicología diferencial a la ciencia natural. El planteamiento central, desde esta perspectiva se ilustra en la figura 6 donde se puede ver como los TR de los sujetos con CI normal son más rápidos que los TR de los sujetos con CI inferior (Jensen, 2005).

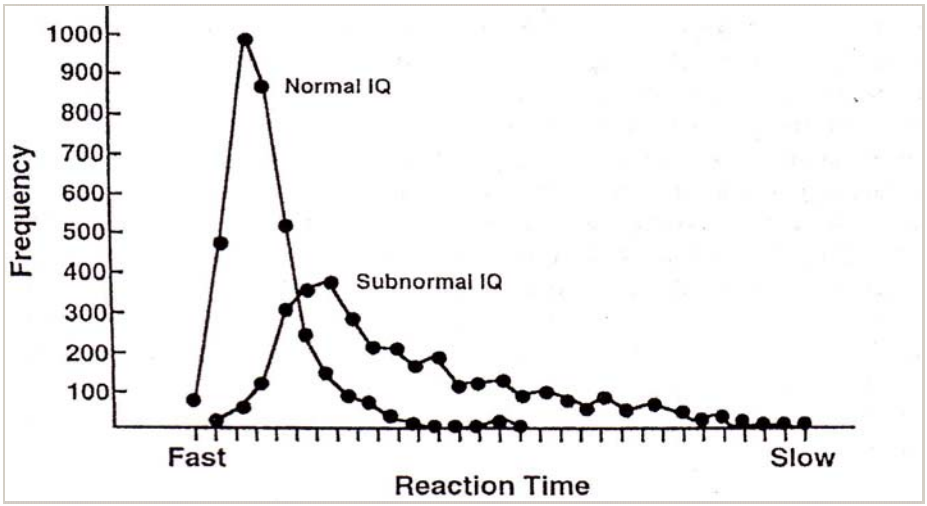


Figura 6. Distribución de los tiempos de reacción de sujetos con CI normal e inferior (Jensen, 2005).

A MODO DE EPÍLOGO

Ha llegado el momento de la recapitulación. Hemos definido la inteligencia como la capacidad de enfrentarse con eficacia a las demandas externas. Pero también como la

eficacia en procesos cognitivos elementales como atención y percepción. Por supuesto las habilidades a la hora de poner en marcha los procesos complejos de razonamiento abstracto, representación mental, solución de problemas o toma de decisiones. También como la habilidad para aprender y, en definitiva, la conducta eficaz en respuesta a los problemas que se nos van presentando. Nos hemos acercado a las teorías más recientes que hablan de que las habilidades cognitivas humanas se pueden agrupar en ocho factores de segundo orden y cerca de noventa capacidades diferentes.

También hemos visto como muchas veces se confunde la inteligencia con lo que miden los test (inteligencia psicométrica). Además, hemos hablado de otras inteligencias no tan 'cognitivas' como la inteligencia social, la inteligencia emocional o la inteligencia interpersonal, que mejores predoctoras del éxito profesional que la inteligencia 'académica'.

Por último, hemos bosquejado unas pequeñas pinceladas de lo que se conoce sobre las psicobiología de la inteligencia.

Y nada mejor para terminar que volver al principio de la exposición y de la humanidad con la descripción que hace Arthur C. Clarke en "2001, Una Odisea Espacial" (1968) sobre los albores de la inteligencia humana:

"…Entre los de su especie Moon-Watcher era casi un gigante. Pasaba un par de centímetros del metro y medio de estatura, y aunque pésimamente alimentado, pesaba unos cincuenta kilos. Su peludo y musculoso cuerpo estaba a mitad de camino entre el del mono y el del hombre, pero su cabeza era mucho más parecida a la del segundo que a la del primero. La frente era deprimida y presentaba protuberancias sobre la cuenca de los ojos, aunque ofrecía inconfundiblemente en sus genes la promesa de humanidad. Al tender su mirada sobre aquel hostil mundo del pleistoceno, había ya algo en ella que sobrepasaba la capacidad de cualquier mono. En sus oscuros y sumisos ojos se reflejaba una alboreante comprensión...los primeros indicios de una inteligencia que posiblemente no se realizaría aun durante años, y no podría tardar en ser extinguida para siempre…"(páginas 14-15 de la edición castellana).

REFERENCIAS BIBLIOGRÁFICAS

Binet, A. y Simon, T. (1905). Méthodes nouvelles pour le diagnostic des niveaux intellectuels des anormaux. *L'Année Psychologique, 11,* 191-124.

Carroll, J.B. (2005). The Three-Stratum Theory of Cognitive Abilities. En D.P. Flanagan y P.L. Harrison (Eds.), *Contemporary Intellectual Assessment* (pp. 69-76). New York: The Guilford Press.

Catell, J.B. (1971). *Abilities: Their Structure, Growth, and Action.* Boston: Houghton-Mifflin.

Chen, J.Q. y Gardner, H. (2005). Assessment Based on Multiple-Intelligences Theory. En D.P. Flanagan y P.L. Harrison (Eds.), *Contemporary Intellectual Assessment* (pp. 77-102). New York: The Guilford Press.

Cianciolo, A.T. y Sternberg, R.J. (2004). *Intelligence: A brief history.* Oxford: Blackwell Publishing.

De Bortoli, M., Barrios, P. y Azpiroz, R. (2002). Relaciones entre los Potenciales Evocados Cognitivos Auditivos y el Test de Matrices Progresivas de Raven. *International Journal of Clinical and Health Psychology*, 2, 327-334.

Dickens, W.T. y Flynn, J.R. (2001). Heritability Estimates Versus Large Environmental Effects: The IQ Paradox Resolved. *Psychological Review, 108*, 346-369.

Extremera, N., Fernández-Berrocal, P., Mestre, J.M. y Guil, R (2005). Medidas de la Inteligencia Emocional. *Revista Latinoamericana de Psicología, 36*, 209-228.

Eysenck, H.J. (1998). *Intelligence: A New Look.* New Brunswick, NJ: Transaction Publishers.

Galton, F. (1883). *Inquiries into the Human Faculty and its Development.* London: McMillan.

Gardner, H. (1983). *Frames of Mind. The Theory of Multiple Intelligences.* New York: Basic Books.

Gardner, H. (1993). *Multiple Intelligences: The Theory in Practice.* New York: Basic Books.

Gardner, H. (1998). Are There Additional Intelligences? The Case for Naturalist, Spiritual, and Existential Intelligences. En J. Cane (Ed.), *Education, Information and Transformation*. Englewood Cliffs, NJ: Prentice-Hall.

Geary, D. C. (2005). The Origin of Mind: Evolution of Brain, Cognition, and General Intelligence . Washington, DC, US: American Psychological Association.

Horn, J.L. y Blankson, N. (2005). Foundations for Better Understanding of Cognitive Abilities. En D.P. Flanagan y P.L. Harrison (Eds.), *Contemporary Intellectual Assessment* (pp. 41-68). New York: The Guilford Press.

Jensen, A.R. (2005). Mental Chronometry and the Unification of Differential Psychology. En R. J. Sternberg & J. E. Pretz (Eds.), *Cognition and Intelligence: Identifying the Mechanisms of the Mind* (pp. 26-50). Cambridge, UK: Cambridge University Press.

McClelland, D.C. (1993). Intelligence is Not the Best Predictor of Job Performance. *Current Directions in Psychological Science, 2*, 5-6

Mayer, J.D. y Salovey, P. (1993). The Intelligence of Emotional Intelligence. *Intelligence, 17*, 433-442.

Mayer, J.D., Salovey, P., Caruso, D.R., y Sitarenios, G. (2001). Emotional Intelligence as a Standard Intelligence. *Emotion, 1*, 232-242.

McGrew, K.S. (2005). The Cattell-Horn-Carroll Theory of Cognitive Abilities: Past, Present, and Future. En D.P. Flanagan y P.L. Harrison (Eds.), *Contemporary Intellectual Assessment* (pp. 136-182). New York: The Guilford Press.

Mingroni, M.A. (2004). The Secular Rise in IQ: Giving Heterosis a Closer Look. *Intelligence, 32*, 65-83.

Myers, D.G. (2004). *Psychology (7th ed.)*. New York: Worth Publishers (Edición Castellana: *Psicología*, Madrid: Ed. Médica Panamericana, 2006).

Nęcka, E. y Orzechowski, J. (2005). High-Order Cognition and Intelligence. En R. J. Sternberg & J. E. Pretz (Eds.), *Cognition and Intelligence: Identifying the Mechanisms of the Mind* (pp. 122-141). Cambridge, UK: Cambridge University Press.

Newman, S.D. y Just, M.A. (2005). The Neural Bases of Intelligence: A Perspective Based on Functional Neuroimaging. En R. J. Sternberg & J. E. Pretz (Eds.),

Cognition and Intelligence: Identifying the Mechanisms of the Mind (pp. 88-103). Cambridge, UK: Cambridge University Press.

Ortiz, S.O. y Ochoa, S.H. (2005). Advances in Cognitive Assessment of Culturally and Linguistically Diverse Individuals. En D.P. Flanagan y P.L. Harrison (Eds.), *Contemporary Intellectual Assessment* (pp. 234-250). New York: The Guilford Press.

Salovey, P. (2001). Applied Emotional Intelligence: Regulating Emotions to Become Healthy, Wealthy and Wise. En J. Ciarrochi, J.P. Forgas y J.D. Mayer (Eds.), *Emotional Intelligence in Everyday Life* (pp. 168-184). Philadelphia: Psychology Press.

Salovey, P. y Grewal, D. (2005). The Science of Emotional Intelligence. *Current Directions of Psychological Science, 14*, 281-285.

Salovey, P. y Mayer, J.D. (1990). Emotional Intelligence. *Imagination, Cognition and Personality, 9*, 185-211.

Salovey, P., Detweiler, J.B., Steward, W.T., y Rothman, A.J. (2000). Emotional States and Physical Health. *American Psychologist, 55*, 110-121.

Salovey, P., Hsee, C. y Mayer, J.D. (1993). Emotional Intelligence and the Regulation of Affect. En D.M. Wegner y J.M. Pennebaker (Eds.), *Handbook of Mental Control* (pp. 258-277). Englewood Cliffs, NJ: Prentice Hall.

Salovey, P., Mayer, J.D., Goldman, S.L., Turvey, C., y Palfai, T.P. (1995). Emotional Attention, Clarity, and Repair: Exploring Emotional Intelligence Using Trait Meta-Mood Scale. En J.W. Pennebaker (Ed.), *Emotion, Disclosure and Health* (pp. 125-154). Washington: APA.

Salovey, P., y Birnbaum, D. (1989). Influence of Mood on Health-Relevant Cognitions. *Journal of Personality and Social Psychology, 57*, 539-551.

Sánchez-Cánovas, J. (1984). *Teorías de la Inteligencia.* Valencia: Promolibro.

Sánchez-Cánovas, J. (1995). Naturaleza de la Inteligencia Humana. En J.M. Latorre (Ed.), *Ciencias Psicosociales Aplicadas* (pp. 287-299). Madrid: Síntesis.

Spearman, C. (1904). General Intelligence: Objectively Determined and Measured. *American Journal of Psychology, 15,* 201-292.

Spearman, C. (1923). *The Nature of Intelligence and the Principle of Cognition.* London: McMillan.

Sternberg, R.J. (1984). Toward a Triarchic Theory of Human Intelligence. *The Behavioral and Brain Sciences, 7*, 269-315.

Sternberg, R.J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence.* New York: Cambridge Univ. Press.

Sternberg, R.J. (1990). *Metaphors of Mind: Conceptions of the Nature of Intelligence.* New York: Cambridge Univ. Press.

Sternberg, R.J. (1996). *Successful Intelligence.* New York: Simon Schuster.

Sternberg, R.J. (2005). The Triarchic Theory of Successful Intelligence. En D.P. Flanagan y P.L. Harrison (Eds.), *Contemporary Intellectual Assessment* (pp. 103-119). New York: The Guilford Press.

Sternberg, R.J. y Kaufman, J.C. (1998). Human Abilities. *Annual Review of Psychology, 49*, 479-502.

Sternberg, R.J. y Wagner, R.K. (1993). The G-ocentric View of Intelligence and Job Performance is Wrong. *Current Directions of Psychological Science, 2,* 1-4.

Terman, L. (1916). *The measurement of intelligence.* Boston: Houghton Mifflin.

Thurstone, L.L. (1938). *Primary Mental Abilities.* Chicago: The University of Chicago Press.
Tortosa, F. (1998). *Una Historia de la Psicología Moderna.* Madrid: McGraw-Hill