RAZONAMIENTO ORDINARIO: MODELOS MENTALES Y MODELOS FORMALES[1]

Jesús M. Larrazabal y Luis A. Pérez Miranda

Dpto. de Lógica y Filosofía de la Ciencia. UPV-EHU

En este artículo se señala la relevancia de la interacción entre los modelos mentales (psicología cognitiva) y los modelos formales (inteligencia artificial) para la construcción de modelos del razonamiento ordinario desde una perspectiva cognitiva. Del estudio comparativo de ambos modelos se derivan algunos problemas ligados a la representación de las actitudes proposicionales y asimismo se sostiene que es importante seguir manteniendo a la lógica, en un sentido amplio, como patrón-guía que nos permita articular y, posiblemente, implementar algunos de los procesos deductivos de razonamiento que han sido previamente acotados en función de nuestros intereses o basándonos en nuestras intuiciones.

Palabras clave: modelos mentales, modelos formales, psicología cognitiva, inteligencia artificial.

En ocasiones, cuando nos hacemos con cierta información a través de las proposiciones del lenguaje natural, somos conscientes de que es posible su sintetización de modo que podamos servirnos de ella más provechosamente. Es decir, reestructuramos la información sin necesidad

---

[1] Este artículo es una versión ampliada de la contribución presentada al ICCS-89. Nuestro agradecimiento a los miembros del seminario de Lógica de la Universidad del País Vasco por los comentarios recibidos. Este trabajo se integra parcialmente en el proyecto de investigación del Departamento de Educación, Universidades e Investigación GB 003.230-0004/88 del Gobierno Vasco. El segundo autor es becario en el programa de investigadores del M.E.C.

de añadir nuevos contenidos semánticos, deducimos consecuencias implícitas o razonamos por defecto, etc. Este tipo de procesos podríamos englobarlos en lo que se ha venido a denominar *Razonamiento Ordinario*. Si bien es cierto que se haya vinculado con otros procesos como el procesamiento de la información y la obtención del conocimiento ligados ambos a la determinación y caracterización de la inteligencia humana, tiene pleno sentido ocuparse del mismo tal y como emergerá a lo largo de este artículo.

Fundamentalmente parece haber dos tipos de aproximaciones de carácter general al razonamiento ordinario, si bien hay además otros enfoques como el razonamiento analógico o los razonamientos ligados a teorías pragmáticas o sociales, que no los trataremos por reducirse usualmente a alguno de los anteriores. Por un lado, los enfoques de orden psicológico que pretenden dar cuenta de este proceso en su sentido más amplio: tanto la *competencia* - qué es lo que la mente computa y por qué - como la *actuación* - los procesos mentales subyacentes; y por otro, algunas de las líneas investigativas cercanas a la Inteligencia Artificial (IA) que buscan la implementación de modelos computacionales del razonamiento, y cuyo patrón de trabajo queda definido básicamente por la búsqueda de eficiencia en cada una de las teorías que puedan derivarse. Un artefacto inteligente debe poseer varias formas de conocimiento y creencias acerca de su mundo, y debe utilizar esa información más completa de modo que le permita tomar decisiones, planificar y llevar a cabo acciones, responder a otros agentes, etc. Son precisamente estos aspectos intencionales, en conexión con la teoría de la acción, los que complementan un estudio puramente teórico del razonamiento ordinario.

El problema técnico para la IA es caracterizar los patrones de razonamiento requeridos y realizarlos computacionalmente (Reiter, 1987). Estos últimos, de concepción claramente formalista, tan sólo recientemente han comenzado a preocuparse de forma seria por construir teorías de carácter ´´completo'' que pudieran abarcar de alguna manera los aspectos de contenido semántico presentes en el lenguaje natural, y que juegan al parecer un papel importante en cada uno de los razonamientos que llevamos diariamente a cabo. En realidad, lo que hay que clarificar es *en qué términos y en qué medida afectan las representaciones de contenido sobre los mecanismos formales de inferencia*. En principio este es uno de los

problemas a resolver para ambas perspectivas. Los psicólogos cognitivos cercanos a la línea de trabajo defendida por Johnson-Laird (1983), principalmente a raíz de la publicación del libro *Mental Models* , propusieron la interesante idea de incorporar los aspectos de contenido semántico a la hora de valorar en su justa medida este tipo de procesos. En general, la interpretación y posterior manipulación deductiva de la información recibida, tanto por un ordenador (base de datos) como por una persona (memoria) tienen como referencia común tanto la coherencia como la no incompatibilidad a la hora de construir un modelo en el que pueda ser recogida toda la información dada. Las condiciones de verdad parecen jugar un papel fundamental en este cometido. No obstante, no queda claro cómo pueden combinarse aspectos puramente formales, o si queremos, conceptos más característicamente lógicos (consistencia, validez, satisfacibilidad, etc.), con los de orden semántico-linguístico (creencias, actitudes, propósitos, etc.). De hecho, hay que establecer una 'diferencia' entre razonamiento teórico y razonamiento práctico en función de la introducción o no en el modelo computacional de las actitudes proposicionales del razonador. Tal y como veremos esta 'diferencia' ayuda a comprender algunos de los malentendidos que se han producido en el área del razonamiento ordinario. El propósito de este trabajo consiste precisamente en la clarificación de esta relación que permanece aún oscura para la ciencia cognitiva.

Lo que resulta explícitamente normativo, esto es, lo que preserva la coherencia discursiva, aparece como un impedimento cuando tratamos de recoger otros aspectos de la comunicación como intenciones o creencias que son de difícil formalización. Todo parece indicar que constantemente manipulamos modelos o estereotipos -*Scripts, Mops* (Schanck 1981, 82); *Frames* (Minsky 1975)- tanto para comprender lo que se nos quiere comunicar como para registrar esa información ya comprendida. También es inmediato que estos modelos son de orden temporal y nunca definitivos, sino que más bien están a expensas de la información subsiguiente que podamos recibir. Actúan como patrones que nos permiten funcionar, especialmente, en nuevas situaciones, no sólo de interacción comunicativa o social, sino de mera comprensión de estados de cosas. Sin embargo, estos modelos deben ser lo suficientemente maleables (o plásticos) de modo que puedan reestructurarse fácilmente y den pie a la introducción de otros aspectos discursivos para que la comunicación sea lo suficientemente

efectiva o nuestra interacción social sea la adecuada, pongamos por caso. Si comprender consiste en establecer la función correcta entre la información almacenada en la memoria y el discurso presente, razonar no consiste sino en extraer conclusiones válidas de esa información de modo que sin pérdida de contenido semántico hagamos abstracción de la forma más sintética posible. En relación a este punto cabe señalar que es perfectamente posible asumir una diferencia entre comprensión y razonamiento en ámbitos discursivos. Podemos haber comprendido, por ejemplo, que Pedro es más alto que Miguel y Miguel es más alto que Juan sin necesidad de derivar que Pedro es más alto que Juan. El punto principal del enfoque semanticista defendido por Johnson-Laird (1983, 86, 88) consiste precisamente en casar razonamiento y comprensión. La comprensión depende de la construcción de modelos mentales y es de hecho similar al proceso de mantener consistente una base de datos. Es necesario llevar a cabo inferencias a partir de la información de entrada, así como revisar la base de datos toda vez que tales conclusiones entren en conflicto con la información subsiguiente.

Siguiendo la línea aristotélico-leibniziana que desemboca en Boole, el razonamiento respondía en especial al uso de determinados formalismos lógicos que de alguna manera presentes en la mente y entrelazados al lenguaje natural, permitían la derivación de conclusiones válidas. Esta perspectiva queda hoy en día representada desde un punto de vista teórico global por Fodor y Phylyshyn (1988) defensores de una Ciencia Cognitiva computacionalmente clásica. En el área específica de la *"Psicología Cognitiva"* este paradigma queda enmarcado en lo que se conoce como *"Lógica de la Mente"* . De hecho, podría establecerse un paralelismo entre la idea Chomskiana (1968) de la posesión de una gramática universal en nuestras mentes *(Competencia),* con la idea de que nos servimos de reglas lógicas (de la lógica simbólica clásica) para derivar conclusiones a partir de ciertas premisas dadas. Esta posición ha sido duramente criticada tanto desde puntos de vista psicológicos como desde posicionamientos más cercanos a la IA : *la lógica simbólica clásica resulta rígida en exceso tanto para la competencia como para la actuación respecto al razonamiento.* Esto explica por qué en psicología cognitiva cuando se ha tratado de buscar un mecanismo o procedimiento explicativo de este tipo de fenómenos se ha procurado configurar modelos de carácter formal, que conservando la

validez en el razonamiento, mantuviera al mismo tiempo un interés declarado hacia las cuestiones interconectadas con el contenido semántico en el discurso ordinario.

En principio la orientación parece la adecuada, pero en realidad lo conseguido, tal y como mostraremos más adelante, no va mucho más allá de la plasmación de algunas de las propiedades pragmático-semánticas de las expresiones linguísticas. Esto es, se ponen de manifiesto ciertos problemas en los modelos computacionales en conexión con el papel que juega el contenido en las deducciones que tanto humanos como máquinas llevamos a cabo. En cualquier caso, reiteramos la idea de que no es sencillo establecer en qué términos podemos combinar los aspectos, digamos, ligados al discurso en el lenguaje con aquellos otros de naturaleza estrictamente lógica. Al mismo tiempo, no resulta fácil imaginar cómo podríamos prescindir de forma definitiva de algún instrumento lógico (*Lógica en los modelos mentales (semánticos); extensiones de la lógica de primer orden: Lógica Modal, Lógica NoMonótona, etc.*) para extraer conclusiones que podamos definir como correctas en función de la información presente en un momento determinado para una base de datos. De este modo, partiremos de los problemas presentes en teorías psicológicas (Johnson-Laird) en torno a la idea de modelo mental. Estos nos darán la pista para comprender en qué sentido las formalizaciones presentes tanto en este tipo de heurísticos como los utilizados por enfoques que utilizan lógicas no-clásicas se encuentran con el impedimento, ya señalado, que constituye la sutileza expresiva y el filón semántico del lenguaje natural.

Obviamente no podemos establecer una relación de perfecta adecuación entre la lógica y el lenguaje natural; tan sólo podemos hacerlo con parte de este último, y es esa parte precisamente la que debemos delimitar. Paralelamente incidiremos en los problemas representacionales derivados de esa relación siempre tan difícil. Este tipo de modelos incorporan información que se generaliza a cualquier razonamiento independientemente de su contenido semántico (paradójicamente, la idea primigenia de Johnson-Laird de trabajar con los contenidos de las deducciones en el momento de llevar a cabo razonamientos no tiene lugar en los modelos por el propuestos). En primer lugar, veremos que en esos modelos el factor formal juega un papel determinante y predominante a la vez. Es decir, una

vez que es recogida cierta información, si queremos obtener deducciones inmediatas a partir de la misma, debemos de servirnos de recursos lógicos para hacerlo. *"Conditio sine qua non"* para mantener la consistencia entre las inferencias realizadas. A nuestro entender los modelos mentales no son sino un tipo de lógica disfrazada. Posteriormente, derivaremos algunos de los problemas a raíz de la asunción de una teoría semejante. Por ejemplo, la teoría del significado propuesta parece sólo funcionar en contextos restringidos donde el dominio de trabajo responde a modelos semántico-formales de orden tarskiano.

Ciertamente, esta crítica también es extensible a las posiciones más en conexión con programas en IA que pretenden explicar este tipo de fenómeno. *¿Cuál es la relación entre los lenguajes mentales y el mundo? En otras palabras, ¿de dónde emerge la semántica? Es interna, externa, de carácter general, dependiente del contexto, etc. ¿Qué papel juega respecto a los procesos deductivos?*

En definitiva, la idea que queremos resaltar es la siguiente: independientemente de que nuestros razonamientos deductivos se ajusten o no al mundo (aunque parece justo apuntar que de alguna manera deben hacerlo cuando llevamos siglos de tradición lógica) como seres inteligentes (nos sirvamos o no de él) tenemos a nuestro alcance un patrón lógico guía común como herramienta de trabajo precisa en la búsqueda de una referencia (framework) coherente para nuestros propósitos, planes y acciones consecuentes. Parece claro que nos hacemos con una parte interesada de la información que nos llega. Esto es, una información en la mayoría de los casos incompleta acerca de un estado de cosas con el que nos debemos manejar. En otras palabras, aquí se está hablando de lo que supuestamente llevaría a cabo un razonador ideal una vez que se ha hecho con cierta información. Debe notarse que cuando hablamos de procesos deductivos en agentes ideales no nos referimos a seres omniscientes al margen de cualquier restricción espacio-temporal, sino al modo en que potencialmente llevaríamos a cabo este tipo de procesos. Y, en algunos casos, a procesos de razonamiento que manipulan mayores recursos computacionales de los que cabe esperarse en los razonadores humanos. En este punto hay que establecer una distinción entre el procesamiento de la información y la posterior manipulación de ésta a través de los mecanismos sintáctico y/o semántico formales de inferencia. Por ejemplo, en IA

generalmente la cuestión no consiste tanto

en ver qué tipos de expresiones relacionamos en función de nuestros intereses, de nuestra capacidad memorística o cosas similares, sino más bien en resaltar los puntos subyacentes a la representación y posterior uso de modo inteligente de la información. En esta disciplina son de sobra conocidas las dificultades para abordar formalmente las cuestiones relativas a creencias, intenciones, actitudes... La cosa es sencilla : los recursos de carácter formal disponibles no resultan suficientes. Es más, problemas de naturaleza incluso más "simple" están todavía por resolver. ¿Cabe modelizar el razonamiento 'deductivo' en ámbitos discursivos en los que las fluctuaciones de información son constantes y a veces impredictibles (Razonamiento Suposicional)? Es decir, lo que nos interesa es el tratamiento de la información deductivamente tanto para procesos de razonamiento práctico como teórico.

No es probable que cada uno de nuestros actos mentales esté constreñido al uso de mecanismos puramente formales o lógicos. Desde esta posición, estaríamos defendiendo una lógica de la mente en su estilo más puro. La idea intuitiva es que en "determinados momentos" utilizamos patrones de conducta lógicos, no sólo cuando de tareas de orden matemático o lógico se trata, sino siempre que pretendamos manipular la información en función de sus consecuencias implícitas. Cuando menos, con esto se quiere decir que nuestras deducciones realizadas son válidas: siendo verdaderas las premisas lo es necesariamente la conclusión. Esta es satisfacible en todos los modelos asociados a las premisas con respecto a un intervalo temporal. Una fórmula satisfacible en un determinado momento por un modelo que recoja toda la información disponible figurando como representante de la misma, puede no serlo en un modelo posteriormente creado en función de la información ulterior. Y es más, estas revisiones de creencia pueden no sólo depender de la información global que cabe examinar lógicamente en cada momento, sino que a su vez pueden verse restringidas por imperativos de orden pragmático ligados a las acciones cotidianas, que tienen su origen causal en las consecuencias obtenidas por el razonador a partir de una parte de la información total disponible.

La idea de reestructuración y síntexis a partir de ciertas premisas también está presente en los modelos mentales de Johnson-Laird. Dadas determinadas premisas, vamos construyendo modelos de las mismas de

modo que no quepa formular nuevos modelos donde la conclusión obtenida sea falsa en alguno de ellos, mientras que las premisas se mantengan verdaderas. En general llevamos a cabo deducciones por defecto, aquí una conclusión determinada puede seguirse siempre que no entre en contradicción con información posterior. En IA se estudian las propiedades de este tipo de razonamientos basados especialmente en lógicas con defectos o similares. Los intereses de este campo de trabajo residen en la manipulación de deducciones en lógicas estrictamente no monótonas o con defectos de ciertas reglas de inferencia (Reiter, Mc Dermott, Doyle) u axiomas de primer o segundo orden con fines de minimización como la Circunscripción (McCarthy) que recaen sobre la representación del conocimiento presente en la base de datos. Desde este planteamiento son importantes las propiedades tanto lógicas como computacionales que podamos derivar. Este asunto quedará desarrollado en un capítulo específico para las lógicas no monótonas y extensiones de las mismas en cuanto a la representación de la información y su posterior razonamiento. Normalmente, estos tipos de razonamiento no son computables al carecer de recursividad en los casos no restringidos. No es sencillo saber cómo regirse en el momento de la elección entre una de las extensiones derivables posibles (a veces incompatibles) de una teoría determinada sin más mecanismos que los puramente lógicos. Aquí vienen a colación las cuestiones referentes a creencias, actitudes, intereses, adecuación, etc. De alguna manera, los enfoques de IA tienden a buscar modelos minimales o teorías de mundos cerrados donde la información que se debería tener para la resolución de una tarea está presente. Esta es una idea ligada tanto a lo que se conoce como el problema del MARCO, como al problema de la PERSISTENCIA (Etherington, 1988). Ambos no son sino distintas caras de la misma moneda. *" Notice that persistence reasoning is different from closed-world reasoning : although the persistences are not entailed by what is known, they are to be assumed true in the absence of contrary information, not false. The difference disappears, though, if we view persistence as a problem of closed-world reasoning about the state of the world itself. This turns out to be a common feature of many of the problems discussed (...) seemingly different approaches turn out to be intimately related if one shifts one's representational framework "* (Etherington, 1988, p.4).

¿Qué tipo de explicaciones pueden darse desde enfoques normativos en cuanto a la idealización del razonamiento que tengan relevancia más allá de los ámbitos puramente formales? En otros términos, ¿ tiene o no tiene sentido hablar de razonamientos supuestamente ideales, en los términos anteriormente indicados, sabiendo que gran parte de las veces, a pesar de tener a nuestro alcance los recursos lógicos suficientes, nos vemos influenciados por constricciones de todo tipo que nos impiden razonar "correctamente"? La cuestión no es tanto que nos veamos obligados a reconocer ciertas reglas lógicas en nuestras mentes, sino más bien a admitir que potencialmente (competencia) podemos servirnos de recursos lógicos al modo de los sistemas axiomáticos (lógicas no monótonas como la lógica autoepistémica, la lógica con defectos...); además de los modelos propuestos por Johnson-Laird, el razonamiento rechazable de Pollock o las lógicas preferenciales (como las de Shoham) en el momento de trabajar deductivamente con información.

1. Inhabilidad de la Lógica Clásica en las Tareas de Razonamiento

En los últimos años, la lógica simbólica clasica, tal y como se ha señalado con anterioridad, ha sido objeto de criticas al no constituir un instrumento adecuado para dar cuenta de la intrincada cuestión de cómo es posible razonar acertadamente en situaciones ordinarias. Hasta la fecha se pensaba que éramos intrínsecamente racionales porque poseíamos una lógica en nuestras mentes. Esta cuestión permitía comprender los errores, falacias, etc. en el razonamiento, no por la falta de principios inferenciales sino por la mala aplicación de las reglas, o por malentendimiento de las premisas, pongamos por caso. No obstante, el argumento presentado por los defensores de la lógica de la mente (mental logic) no es falsable empíricamente. *"If experimental circumstancies which would conclusively falsify a hypothesis cannot be specified, then it is not a scientific hypothesis; and if some of the assertions of a discipline are not falsifiable, then it is not a science "* (Malcolm, 1990, p.5). Es más, para muchos psicólogos cognitivos el razonamiento ordinario no es de orden estrictamente determinista, sino que cuando llevamos a cabo razonamientos prácticos o intuitivos nos preocupamos más de ser realistas que de ser consistentes. En estos casos, nuestro objetivo consistiría en elaborar modelos mentales interpretativos de la realidad teniendo en cuenta las

relaciones causales, la plausibilidad, etc. conceptos que en principio figurarían al margen de la lógica. Ahora bien, a partir de un "corpus" de conocimiento específico (independientemente del modo en que hayamos conseguido ese conocimiento), podemos obtener deductivamente una reorganización o reestructuración del conocimiento que tan sólo se encontraba implícito en las premisas, aunque no suponga una ampliación de la información semántica. La cuestión es establecer con claridad cómo nos servimos de este tipo de mecanismos de inferencia cuando queremos preservar la coherencia discursiva, cuando queremos usar cierta información de modo inteligente sin caer en contradicciones (una base de datos por ejemplo), cuando queremos sintetizar información.

1.1 Crítica Extendida a los Sistemas de Reglas de Inferencia en General

La crítica más severa recibida por los posicionamientos formalistas viene de la mano de los psicólogos cognitivos y recae sobre la asunción al parecer indebida de que el mecanismo de deducción basado en las reglas de inferencia (*modus ponens, resolución* en todas sus variantes), no opera sobre el significado de los enunciados sino sobre su forma abstracta. Según Johnson-Laird y sus seguidores, tales formalismos prescinden del significado de las expresiones a la hora de derivar conclusiones, cuando de hecho el contenido jugaría un papel definitivo en el razonamiento, y por tanto, su análisis se situaría inadecuadamente en un nivel exclusivamente sintáctico. Además señalan que este enfoque se enfrenta con la dificultad para determinar qué conclusiones particulares válidas son inferidas espontáneamente. Como sabemos, una lógica completa nos permitiría ratificar toda conclusión que se derivase de un conjunto de premisas, aunque de hecho cualquier conjunto de premisas implique un número infinito de diferentes conclusiones válidas. De este modo, quedarían admitidas todo tipo de reiteraciones y repeticiones que podrían suponer una pérdida de la información semántica, que en principio no parece lícita en el razonamiento ordinario. Por ejemplo, si obtuviéramos una conclusión q a partir de una regla de inferencia que preservara la validez nada nos impediría añadir a la misma como disyunción cualquier enunciado p. Ahora bien, no parece probable que cualquiera de nosotros durante un razonamiento corriente lleváramos a cabo una conclusión semejante: pvqv...vq sucesivamente, independientemente del supuesto de que la

conclusión fuera lógicamente válida. Lo más inmediato sería debilitar en alguna medida la lógica utilizada de modo que pudiera acoplarse a este tipo de razonamiento.

No obstante, la propuesta de los psicólogos cognitivos consiste en la utilización de los *modelos mentales,* que sin necesidad del uso de la lógica (Johnson-Laird 1986), permitirían explicar este tipo de razonamientos en un sentido general, salvo para razonamientos en los que intervinieran extensiones infinitas de elementos (lógica, matemática,...), de modo que quedaran incorporadas definitivamente aquellas propiedades discursivas de carácter semántico. Como veremos, además de los inconvenientes de esta propuesta en cuanto a la teoría del significado, la plausibilidad, etc. su fundamento básico de tener en cuenta los contenidos de las proposiciones no tiene lugar en este tipo de modelos. En realidad, configuran un proceso de orden lógico formal. Los contenidos sin duda ayudan a la estructuración de la representación de un ámbito, o mundo (posible o natural). Los factores de capacidad memorística, o de interés específico por una parte concreta de la información, preservan un tipo de información para desechar otra. Sin embargo, una vez que tenemos presente una determinada representación de aquélla, nuestro razonar podría discurrir de un modo más cercano a la lógica de lo que en principio parecería suponerse. Sería interesante revisar los comentarios al respecto por parte de Stenning(1990) o Lee(1987), donde se puntualiza que los modelos mentales no constituyen más que un tipo de razonamiento de orden lógico para pesadumbre de algunos psicólogos. De hecho, se han encontrado relaciones de isomorfismo entre las deducciones llevadas a cabo en los modelos mentales tanto con los *Círculos de Euler*, como con los procesos de *Deducción Natural de Gentzen.*

Otros enfoques también criticados por Johnson-Laird son los conocidos sistemas expertos y las reglas de inferencia de contenido específico; tales reglas son condicionales que pueden ser representados dentro del marco de un sistema de producción, y los programas de cómputo de este tipo han sido desarrollados en orden a recoger parte de la pericia humana. Sus reglas de inferencia tienen un conocimiento específico extraído de los expertos humanos. Los sistemas expertos resultantes proporcionan una estimable ayuda en terrenos de lo más amplios. Un programa tal funciona a través de reglas para dar una solución a problemas particulares propuestos

por la persona que usa el programa. Los programas para el razonamiento en sistemas expertos varían considerablemente. Algunos usan reglas que funcionan ascendente o descendentemente; otros programas usan estimaciones de las probabilidades de las hipótesis, y de la probabilidad de las observaciones particulares referentes a cada hipótesis dada *(aproximación bayesiana)*. Sin embargo, independientemente del método, un sistema es sólo tan bueno como el conocimiento que incorpore, y el asunto de formular conocimiento explícitamente es dificultoso porque gran parte de él no es inmediatamente adecuado para introspección. Los sistemas expertos fallan en cuanto teoría completa del razonamiento. Puesto que no proporcionan un mecanismo para la habilidad inferencial en general. Después de todo, la gente puede llevar a cabo inferencias válidas en dominios que resultan poco o nada familiares. En opinión de Johnson-Laird lo que se necesita es una habilidad general acompañada de sensibilidad con el contenido.

2. Contextualización del Razonamiento Ordinario

Normalmente, cuando nos enfrentamos a un problema social o intelectual, el propósito puede ser preciso, es decir, estaríamos en disposición de circunscribir o no nuestra información a la hora de dar cuenta de nuestro problema. No obstante, en escasas ocasiones disponemos de un procedimiento establecido para razonar a partir de un conjunto de premisas llegando a una conclusión, tal y como ocurre en el ejemplo siguiente. Ej.: Si pasas los jueves a la tarde por el *"Deacon Pub"* , probablemente verás a Amina. Esta proposición retenida en la memoria podría tener un interés evidente, por ejemplo, si quisieramos encontrarnos con Amina y esa idea nos rondara la cabeza precisamente un jueves por la tarde. En cualquier caso, nuestra conducta podría guiarse de alguna manera en función de esta información con independencia del tipo de resultados que pudieran desprenderse de su uso. En general, en un razonamiento conectamos las premisas con la conclusión. En el caso de la deducción, esta relación es la de validez, es decir, que la conclusión debe ser verdadera supuesto que las premisas lo son. En este punto cabe matizar que la validez no expresa que las premisas sean vedaderas, sino que si lo son, entonces lo es la conclusión (aquí por el momento nos olvidamos de la especial caracterización de los conceptos semánticos cuando introducimos el

concepto de no-monotonía en nuestra lógica clásica). Los límites de los distintos tipos de razonamientos pueden establecerse en función del concepto de información semántica Johnson-Laird (1988). Cuantos más posibles estados de cosas elimine una proposición de cierta consideración, mayor será la información semántica que contenga. Si nos fijamos ésta es también la idea presente en los modelos minimales de la Circunscripción propuesta por McCarthy(1980). Una inferencia será válida cuando se mantenga la cantidad de información semántica presente en las premisas, es decir, cuando no haya pérdida de información. Sin embargo, si la conclusión elimina estados de cosas adicionales, entonces no será válida.

Cuando se nos presentan ciertas premisas de la vida cotidiana, como por ejemplo: *el cadáver apareció tendido sobre el suelo en un cine madrileño. El sospechoso viajaba en un expreso camino de Bilbao cuando ocurrió el trágico suceso*. En principio, las posibles situaciones que se pueden inferir están en función del conocimiento general que tengamos en torno al suceso, lo cual nos ayudará a configurar extensiones consistentes sobre la base de la información disponible, seguramente incompleta, que se ajustará en mayor o menor medida a la realidad. Sin duda, en este tipo de deducciones se aprecia un tipo de *similaridad estructural* con las teorías con defectos (lógicas con defectos). Partiendo de una teoría con defectos T= < W, D > compuesta por un conjunto W de expresiones representativas de un mundo, junto con un conjunto de reglas con defectos D aplicables según la información disponible en ese momento, obtenemos extensiones de la teoría. Una vez que el problema se sitúa en punto a la derivación de algunas conclusiones válidas puede darse el caso de que resulten totalmente triviales o redundantes. De alguna manera, queda patente el hecho de que en los razonamientos ordinarios no nos podemos permitir el lujo de extraer conclusiones que supongan una pérdida de la información semántica.

3. Idea de Modelo Mental

El procedimiento "semántico" de los modelos mentales está constituido por las siguientes etapas: interpretar las premisas construyendo un modelo basado en sus condiciones de verdad, formular una conclusión informativa, y confrontar la conclusión buscando diferentes modelos para las premisas. Concretamente nos serviremos del caso del razonamiento silogístico para

su ejemplificación. Supongamos que se nos da la premisa Todo B es A. En primer lugar (i) deberíamos *construir un modelo mental de la misma:*

b=a
b=a
(a)
(a)

El número de (elementos) "tokens" correspondientes a B y A es arbitrario, y los "items" entre paréntesis representan la posible existencia de a's que no son b's. Cada una de las premisas requeriría la existencia de un solo modelo mental simple.

Si introducimos una nueva premisa, Ningún B es C, por ejemplo, deberíamos *añadir la información de esta última al modelo mental de la primera (ii), teniendo en cuenta los diferentes modos en que esto puede llevarse a cabo.* En todos los casos no debe pasar desapercibido el principio semántico fundamental subyacente a la deducción válida: una inferencia es válida si no hay modo de interpretar las premisas tal que resulten consistentes con la negación de la conclusión. Ambas premisas combinadas dan lugar a tres modelos diferentes;

1. c Ninguno de los C es A o su conversa
   c
   .....
   b=a
   b=a
2. c
   c=a
   .....
   b=a
   b=a

(a) Falsea la conclusión anterior y sugiere, no obstante, que Algunos de los C no son A, o conversamente, Alguno de los A no son C.

3. Nuevo contraejemplo:

c=a
c=a
..... Algún A no es C.
b=a
b=a
(a)

Finalmente, *(iii) debemos enmarcar una conclusión para expresar la relación, si existe alguna, entre los términos "finales" que aparezcan en todos los modelos de las premisas.* En este caso sería Algún A no es C. Un término final ocurre sólo en alguna de las premisas, a diferencia del término medio que ocurre en ambas premisas. Si no existe una relación semejante entre los términos finales, las únicas conclusiones válidas que pueden derivarse son triviales (conjunciones o disyunciones).

Los ejemplos propuestos por Johnson-Laird son ilustrativos en el caso del silogismo, pero no aclaran del todo la cuestión general para razonamientos más complejos. El hecho de que nos veamos afectados por el contenido de un problema cuando razonamos, lo único que sugiere, al menos planteado de esta manera, es que en esos casos estamos probablemente derivando conclusiones indebidamente, o siguiendo caminos menos prácticos y de más "coste" mental. Parece evidente que el transfondo conceptual desarrolla un rol importante en lo que se refiere a la relación y uso de la información, siendo otra cuestión la de que nos sirvamos adecuadamente o no de ciertos mecanismos potencialmente a nuestro alcance para reestructurar inteligentemente nuestra información. Fijémonos en el siguiente ejemplo propuesto por Johnson-Laird: *Sean dadas las cartas expuestas a continuación: A, B, 2, 3 con letras por un lado y números por otro.* La cuestión es decidir qué cartas necesitas dar la vuelta para determinar la verdad o falsedad de la regla: *"Si no hay una vocal por un lado de la carta, entonces hay un número par por la otra".* Curiosamente, la gente se inclina por descubrir A o 2, mientras que la respuesta correcta sería 3. El hecho de que gran parte de la gente efectúe una elección no adecuada, no indica solamente que nos veamos afectados por el contenido de las proposiciones, sino que además, y esto es lo que no señala Johnson-Laird, podríamos llevar a cabo una elección más inteligente

en el caso de que estuviéramos acostumbrados o entrenados con este tipo de razonamientos. Lo que queremos dar a entender, no es que las personas habitualmente hagan un *uso estricto de principios lógicos* cuando de derivar conclusiones se trata, sino más bien se pretende mostrar la necesidad de algún tipo de lógica como *patrón guía* cuando trabajamos con información. Hay muchos factores restrictivos que rodean a un razonamiento : memorísticos, de creencia, propósitos, temporales, contextuales, etc. Ahora bien, la referencia última que se desprende de cada uno de ellos tiene o puede tener un fundamento lógico como base; en otro caso, ¿qué alternativas nos quedan?, ¿qué otros mecanismos podrían sustituir a la lógica como teoría de la competencia? Nuestro planteamiento continúa la línea del principio de "Razonamiento Minimal" propuesto por Cherniak(1986) *a minimal rationality (agent) must have some, but not ideal, logical ability -that is, he must satisfy the minimal inference condition: The agent would make some, but not necessarily all, sound inferences from his belief set that are apparently appropiate (p.9).*

4. Teoría del Significado

Asunciones principales de la teoría del significado de los modelos mentales: (i) Propiedades: las proposiciones pueden referirse al mundo; no obstante, los seres humanos no aprehenden el mundo directamente; sólo poseen una representación interna de él, porque la percepción es la construcción de un modelo del mundo. Las proposiciones pueden también referirse a mundos imaginarios o hipotéticos. Una proposición puede ser falsa para un modelo dado que las otras premisas que intervienen en el razonamiento son verdaderas con respecto a él y la introducción de aquélla provocaría una contradicción entre las proposiciones de cara a la configuración del modelo. Los seres humanos pueden construir modelos mentales mediante actos de imaginación y pueden relacionar proposiciones con tales modelos Johnson-Laird(1983).

(ii) Dificultades: como vemos, la relación entre premisas y conclusión se traslada a un lenguaje mental. Sin embargo, esta teoría debería aclarar determinadas cuestiones que no quedan bien establecidas. ¿Cómo se relacionan las representaciones proposicionales con los modelos imaginarios o reales?, ¿cuál es la relación entre modelos y mundo?. En algún sentido,

si a partir de un conjunto de proposiciones podemos derivar o inferir varios modelos, parece inevitable desde esta teoría que tendremos que comparar los modelos con el mundo para su verificación, pero además habría que explicar en qué medida son plausibles y por qué. Pero si no aprehendemos el mundo directamente, ¿cómo vamos a establecer funciones entre el modelo y el mundo?. Además, no hay argumentos suficientes para garantizar que las condiciones de verdad intervengan debidamente en los modelos. No se vé cómo aquellos aspectos discursivos presentes en un *"marco* " se incorporan en el modelo, añadiendo plausibilidad a la coherencia del mismo. En resumidas cuentas, el traslado del problema del significado a un lenguaje mentalista no muy distante del lenguaje del pensamiento fodoriano, no clarifica la cuestión de la representación del conocimiento y su posterior utilización en los procesos de razonamiento. Lo más destacable de este análisis es el hecho de que la configuración en un modelo mental de contenido y forma es posible, en tanto que seamos capaces de establecer una distinción previa: el contenido afecta en cuanto a la predisposición que tenemos para trabajar con una u otra información, en cuanto a la posibilidad de que ciertas proposiciones sean compatibles o no dentro de un mismo modelo, pero no en cuanto a la compatibilidad misma (o nociones similares de carácter lógico) o a las derivaciones posteriores que podríamos llevar a cabo una vez que hemos interpretado adecuadamente la información.

Es decir, una vez que nos servimos de heurísticos o probadores de teoremas similares a los modelos mentales, el patrón que seguimos presenta una naturaleza definidamente lógica. Los argumentos en contra de la lógica no clásica (además de sus limitaciones expresivas) suelen provenir del hecho de que resulta intratable desde un punto de vista computacional (si de la lógica de primer orden se trata, o de las lógicas extendidas con defectos; la primera porque no es recursiva, y por tanto no decidible, mientras que las segundas no son tan siquiera semi-decidibles -ver Reiter(1987). La aplicación de heurísticos se muestra como la mejor solución, y al parecer, si nos servimos del ejemplo de los modelos mentales, el establecimiento de una conclusión es definitivo debido a que manipulamos una cantidad reducida de información que podría ponerse en entredicho con la llegada de nueva información inconsistente con la primera. De alguna forma las capacidades o recursos humanos para el

razonamiento están limitadas y, por tanto, las posibilidades de obtención de las extensiones "válidas" derivables a partir de un conjunto de proposiciones también lo están. Lo que se pretende señalar es que la intratabilidad de la lógica en un sentido global, no afectaría en cuanto a la búsqueda de modelos de índole lógica para el razonamiento ordinario, dado que tan sólo deberíamos servirnos de parte de sus recursos y no de su totalidad Levesque(1988). En realidad está por definir no sólo cómo podemos implementar la lógica, sino también queda por clarificar de qué lógica se trata Fodor & Pylyshyn(1988). Como consecuencia de lo dicho podríamos asumir las palabras de estos autores: ...*"finite performance is typically a result of interaction of an unbounded competence with resource constraints"* (Ibid, p.34). De alguna manera aquí se recoge la idea primigenia de Chomsky(1.968) respecto a la competencia linguística. Algunas cuestiones relativas al significado son resaltadas en el apartado siguiente dedicado al problema de la representación.

5. Problema de la Representación

Usualmente para hacer más manipulable el lenguaje natural se ha tendido a desambiguarlo, en parte, siguiendo las técnicas de traslado de ese lenguaje corriente a un lenguaje formalizado que permita su manipulación efectiva. Más concretamente, la lógica clásica ha servido en parte de apoyo a la representación y posterior uso del conocimiento. Ejemplos de ello los tenemos en la *logicización* del lenguaje natural con el fin de configurar una *gramática universal* por parte de Montague(1974); o en *Principia Mathematica* donde Russell & Whitehead buscaron un cálculo lógico de modo que desambiguando las cuestiones que hemos denominado de contenido pudieran dar solución, entre otros, al problema de la deducción. En la misma línea Boole asumió la relevancia de la lógica para todo lenguaje cotidiano que quisiera mantener unos criterios de minimalidad en cuanto al orden y comprensión discursivos. Este consideraba ineludible el uso de este lenguaje formal, asignándole propiedades universalistas. Un análisis en regla de las expresiones linguísticas usuales, es decir, un escrutamiento de las "leyes" que rigen el lenguaje natural, nos permitiría ir obteniendo progresivamente todas y cada una de las reglas que cabría derivar en todo lenguaje universal. En su opinión la lógica no constituía

sino una de las características esenciales presentes en los lenguajes naturales. Más específicamente, la lógica configuraría la denotación formalizada de ciertos principios o reglas (psicológicas) presentes en la mente. De alguna manera las ideas booleanas nos sugieren que junto a la semántica del lenguaje natural, la lógica aparecería encubierta en el propio lenguaje. Tanto su rigor como sus necesidades pragmáticas serían obvias una vez de haber analizado seriamente aquél. ¿Por qué una idea tan intuitiva se ha desechado en apariencia por parte de algunos psicólogos cognitivos?. Está claro que la lógica nos ofrece un medio representacional único, un lenguaje formalizado bien estructurado, con unos mecanismos inferenciales sintáctico-semánticos de los cuales conocemos sus propiedades claramente expresadas en los teoremas que la metalógica nos ofrece. Por qué entonces trasladar el problema de la representación al propio lenguaje natural y no a un lenguaje formal bien establecido que diera cuenta de parte de las características de aquél. Como anotamos más arriba, el intento de desambiguación linguística de las expresiones discursivas sirviéndonos de un lenguaje formal con una semántica extensional tarskiana, evita por un lado el problema del contenido, contexto, etc. (quizás habría que decir elude), pero por otro pierde parte de las propiedades primigenias básicas de la comunicación y el discurso. Está claro que ninguna lógica conocida puede dar cuenta de modo suficiente del problema de la representación de las proposiciones del lenguaje natural _problemas de cuantificación-determinación, contenido, contexto..., no son completamente expresables en los lenguajes formales. Esto es, no es posible su traslado total a expresiones lógicas.

Implícitamente se están rechazando las intuiciones precedentes en positivistas lógicos como Russell. El lenguaje natural al ser ambiguo dificulta su manipulación mecánica, de ahí su necesidad de formalización. Sin embargo, esta última genera otro tipo de problemas al olvidar parte de las características fundamentales de orden linguístico. A nuestro juicio, una salida con garantías de este *"impasse"* podría constituirlo el establecimiento de los límites entre un tipo de lenguaje y otro. Más específicamente, los procesos deductivos que estamos estudiando podrían clarificar la cuestión en los términos siguientes: la deducción se ha considerado desde siempre como un proceso de orden lógico, donde se asumían ciertas propiedades metalógicas sobre la idea de validez en el

transcurso de la misma. *La cuestión consiste en establecer la continuidad o no de este concepto de deducción cuando nos centramos en ámbitos que trascienden los límites marcados por la lógica.* Es decir,¿cabe formular la deducción, y en consecuencia el problema de la representación desde lenguajes formales (lógicos o similares), sin olvidarnos de las cuestiones pragmático-semánticas que implíci-tamente acompañan a las proposiciones linguís-ticas? La respuesta sería afirmativa desde el punto de vista de la psicología computacional clásica. No obstante este problema en parte representacional toma otras vertientes en términos conceptuales para algunas de las posiciones mantenidas en psicología cognitiva.

La lógica clásica nos proporciona, como hemos mencionado, unos mecanismos sintáctico-semán-ticos de inferencia. La semántica presenta un carácter extensional que parece ser suficiente para las tareas de deducción. Pero si conside-ramos que el uso de las propias reglas de inferencia tiene lugar solamente teniendo en cuenta las cuestiones de contenido, o lo que es lo mismo, que el contenido daría pie a la configuración tanto de lo que debemos representar como de las reglas inferenciales a utilizar, entonces habrá que poner en duda entre otros la aplicabilidad del concepto de monotonía. Otro punto a destacar es que la información de partida en los sistemas de conocimiento y bases de datos es la mayoría de las veces incompleta, representa tan sólo parte del conocimiento adecuado acerca de un mundo; por tanto razonamos a sabiendas de que nuestras conclusiones no son definitivas sino de carácter provisional. A pesar de la incompletud de nuestras teorías no podemos concluir que las proposiciones inferidas a partir de las mismas carezcan de valor informativo y no deban ser asumidas. De hecho cotidianamente debemos tomar decisiones en las que asumimos o damos por supuesto ciertos hechos que no podemos garantizar. *"Assumptions made to deal with the various forms of incomplete information cannot be sound, in the usual sense of never leading from true premises to false conclusions "* (Etherington 1988,p.4). Lo que necesitamos es una *justificación* de nuestras asunciones o suposiciones *"principles which allow gaps in one's knowledge to be filled and which guarantee that -most of the time- this assumptions will not lead too wildly astray"* (Ibid,p.5).

En IA se han buscado soluciones que van desde la *Asunción del Mundo* Cerrado (Closed World Assumption) hasta las distintas formulaciones de las *Lógicas No-monótonas*. En el razonamiento ordinario este tipo de

asunciones implícitas la mayor parte de las veces es de carácter inconsciente, es decir, damos por supuesto que una cantidad innumerable de situaciones permanecen invariables cuando actuamos o razonamos sobre ellas. Son asumidas defectivamente como verdaderas mientras no haya nada que haga suponer lo contrario. Para lo que nos interesa, por el momento sería suficiente retener en la memoria que, aunque los posibles modelos computacionales tendrían que representar de algún modo todo este tipo de invariancias para una teoría de la acción sobre un ámbito o mundo concreto (ver Shoham(1.988)), éstas no son necesarias, a nuestro juicio, para una explicación cabal de los procesos deductivos como tales. Más bien hay que tomar en consideración la cantidad de información disponible en un momento determinado y analizar los mecanismos inferenciales adecuados de los que podríamos echar mano cuando tratamos de obtener extensiones consistentes derivables de aquélla. Sabemos que la información es incompleta, pero a riesgo de deducir conclusiones probablemente no definitivas (inconsistentes entre sí) hacemos uso de ella cuando creemos nos resulta provechoso. En definitiva, estamos justificando la autonomía de un tratamiento estrictamente teórico del razonamiento ordinario que, por supuesto, podría ser complementado desde el punto de vista de una teoría general de la racionalidad, con la introducción del razonamiento práctico que se encargaría de explicar el papel de las actitudes proposicionales. Un problema serio se plantea cuando en la información inicial no sólo hay asunciones implícitas, digamos, en torno al mundo anteriormente descritas, sino que, y esto complica mucho más las cosas, esa información contiene información implícita *anidada* más fuerte que forma parte de las convenciones del lenguaje. En muchos casos no es necesario indicar explícitamente todo lo que se quiere comunicar, sino que el lenguaje mismo nos permite captarlo convencionalmente por defecto. En estos términos, si no nos cercioramos del papel que juegan las asunciones podríamos caer en el error de desechar los mecanismos formales de inferencia sin justificación previa. "*... Inferences are generated both from the sentences making up a task and from the sentence and assumptions introduced by reasoners themselves though the process of encoding and understanding. If then implicit assumptions are not made explicit to an observer, they can change the reasoning process in such a way that the resulting inferences might look incorrect* "(G.Hagert & Y.Waern(1.986), pp. 94-95)

En resumidas cuentas hemos establecido una distinción entre las asunciones implícitas con respecto al "mundo" y aquellas que forman parte del propio lenguaje. Sirvámonos de algunos ejemplos: si hablamos acerca de los objetos que están sobre nuestra mesa de trabajo, damos por supuesto que sus propiedades intrínsecas de color, tamaño, forma, textura, etc. no van a desaparecer durante el intervalo de tiempo en que nos sirvan como referencia a lo largo de este razonamiento, pongamos por caso; de algún modo, esto es distinto en cuanto a *tipicidad* a una asunción del tipo: *"tener humo a la vista apunta la proximidad de fuego"*. A nuestro juicio, si las primeras ya resultan ser lo suficientemente problemáticas para los programas de IA, estas últimas colman el vaso al introducir en nuestros razonamientos cuestiones relacionadas con la intencionalidad ligadas al contenido semántico del discurso que difícilmente pueden ser recogidas (representadas) de forma completamente satisfactoria por modelos formales. Como diría H. Putnam(1988) (refiriéndose al funcionalismo), *una descripción computacional completa de "prueba", "confirmación"," sinonimia", etc, será siempre una imposibilidad.* Es decir, el intento de examinar (y por tanto de representar) conceptos semánticos como el "significado" o la referencia fracasa por la misma razón que fracasa el intento de examinar la razón misma: la razón puede trascender aquello que examina.

6. Problema de la Representación en las Lógicas No-Monótonas

El papel de la lógica formal clásica en la representación del conocimiento ha sido duramente contestado desde el principio debido en parte a problemas que tienen que ver con la historia y propósitos de la representación del conocimiento (RC) y sus diferencias respecto a los de la lógica simbólica. Después de las aportaciones leibnizianas, el siguiente gran empuje en lógica formal provino tanto de los trabajos de Boole como de Frege quienes junto con Russell y Peano dieron a la lógica parte de la pujanza de la que hoy en día disfruta en un gran número de campos científicos. Como es de todos conocido el propósito de estos trabajos de Frege en adelante fue dar un fundamento firme a la matemática. La principal aplicación de la lógica simbólica clásica ha sido el análisis de las teorías formales de conjuntos y números (W&M.Kneale (1.961)).

Los propósitos de la RC, no obstante, fueron bastante diferentes, y quizás mucho más en la línea con el sueño leibniziano (Brachman & Levesque(1.984). Los esquemas de RC fueron usados tanto para representar el contenido semántico de los conceptos del lenguaje natural (redes semánticas, estereotipos, esquemas, etc...) , como para representar modelos memorísticos psicológicamente plausibles (Scripts, Frames, Mops,...). En ningún caso existía una clara relación con los lenguajes formales de ningún tipo. Sin embargo, gradualmente empezaron a usarse esquemas para la RC como un modo muy flexible y modular de representar los hechos que un sistema necesitaba conocer para manejarse inteligentemente en un medio complejo. Este punto de vista de la RC que envuelve representaciones proposicionales de las creencias de un sistema, junto con la línea de la Hipótesis de la Representación del Conocimiento (Smith(1.982)) han llegado lentamente a dominar el campo. Lo que nos dice la Hipótesis de la Representación del Conocimiento es que si encontramos a un agente cuya conducta exhiba conocimiento de algún tipo, entonces si indagáramos en el interior de su cerebro deberíamos encontrar codificaciones simbólicas directas de ese conocimiento. No sólo deberíamos reconocer esto como el conocimiento codificado, sino que deberíamos también verlo como causa de la conducta del agente. Más claramente, si encontramos a un agente que actúa como si conociera algo, entonces una forma codificada de ese conocimiento será un ingrediente esencial en la causación de su conducta. Tal y como señalan los autores arriba citados, se puede decir que hoy en día existe un consenso emergente acerca de algunos de los asuntos que relacionan RC y lógica, al menos entre quienes adoptan la perspectiva de otorgar cierta aplicabilidad a la lógica en RC, sin por ello tener que compartir necesariamente las tesis logicistas fuertes (Nilsson 1991). (Esta hipótesis es la defendida generalmente por las aproximaciones computacionales clásicas en cuanto a la representación del conocimiento. A este respecto parece importante apuntar, aunque con algunos matices diferentes, la diferencia establecida por Dennet entre los niveles de implementación y descripción. Es perfectamente plausible creer en la fuerza explicativa de nuestras teorías representacionales explícitas para el razonamiento, sin por ello tener que admitir que exista una perfecta adecuación entre esta teoría y la forma en que realmente el cerebro pueda llevar a cabo el proceso. Incluso en el caso de que hayamos encontrado una

implementación para nuestra teoría, podemos establecer una diferencia entre los niveles de explicación de alto y bajo nivel. Desde luego, no parece plausible asumir que un razonador como OSCAR Pollock(1991), pongamos por caso, en realidad simule al cerebro en el caso concreto del Razonamiento Rechazable (Defeasible Reasoning). No obstante, para entender el punto en su globalidad hay que tener en cuenta que las teorías del razonamiento así definidas no se reducen a la caracterización de los modelos psicológicos humanos, sino que asumiendo la existencia de este tipo de razonamiento 'reevaluable', tratan de establecer sus propiedades fundamentales dando por válido cualquier medio físico en que esta teoría pudiera ser implementable, con independencia de la limitación de recursos que pudiera establecerse entre un razonador humano y un razonador inteligente artificial. En resumen, las limitaciones de recursos en los humanos no resultan ser un impedimento para el desarrollo de una teoría que podría tener aplicaciones más amplias si el medio físico utilizado tuviera a su vez recursos computacionales más extensos. Por tanto, es razonable asumir que la teoría de la racionalidad se extienda más allá de las propias limitaciones humanas).

Por otra parte, sin una especificación concreta acerca del significado de las expresiones (sin una convención notacional), lo que podría ser implicado por una proposición en un lenguaje concreto resultaría del todo ambiguo, y las comparaciones con otros sistemas notacionales serían imposibles de llevar a cabo. En estos casos se habla de significado independiente de las expresiones, de lenguaje desambiguado o neutro en cuanto a connotaciones, de modo que se pueda llegar a establecer que las conclusiones obtenidas en un programa determinado son correctas o completas. Como resultado de lo anterior el punto de vista emergente es que conviene más recurrir, quizás, a lenguajes lógicos no-estándar en su sintaxis, semántica, y uso, que a otros esquemas representacionales. Es algo erróneo, sin embargo, pensar en los lenguajes para la RC sólo en términos del lenguaje representacional. *Puesto que el papel de un sistema de representación es manipular creencias expresadas en el lenguaje, esto es mucho más que la implementación de un cálculo lógico.* En particular, nuevamente siguiendo la hipótesis de la representación del conocimiento, las estructuras simbólicas en una base de conocimiento causan que el sistema funcione de uno u otro modo. Como tal hay una extensión temporal

en las estructuras de RC que no es ciertamente una propiedad de los lenguajes formales. Hay un número de cuestiones generales inducidas por este aspecto de la manipulación de la representación del conocimiento. El más notorio quizás, es que el *Razonamiento* será necesario para determinar lo que está implícito en las creencias explícitas que tiene un sistema (Brachman & Levesque, Ibid.). El grado de incompletud de estas creencias determinará en gran medida la dificultad de dar con sus implicaciones. Además, dependiendo de la extensión de su incompletud, será a menudo necesario para un sistema hacer asunciones o usar defectos para completar sus creencias y entonces revisar estas conclusiones no-monótonamente en función de la nueva información que es adquirida del mundo. La lógica debería proporcionar para los sistemas de RC un claro enunciado de en qué sentido u otro deberían ser las consecuencias lógicas de un razonamiento.

Acerca de la asunción de estos mismos autores, según la cual, -la lógica por sí misma no puede y no debería intentar proporcionar especificaciones de cómo este tipo de revisión racional de la creencia debería tener lugar-, cabe hacer la matización siguiente : al introducir cambios estructurales puntuales en las distintas lógicas (operadores modales, de creencia, conocimiento, etc.) con el fin de ampliar la semántica formal de modo que tenga cabida el tratamiento de la creencia, la idea subyacente no es la de analizar la motivación (intencionalidad) que mueve a un agente a llevar a cabo su propósito como parte de un plan que tiene su origen causal en sus creencias. Al contrario, dejando al lado las cuestiones teleológicas, se supone que la intencionalidad del agente viene "dada" y, por tanto, es del todo plausible asumir la posibilidad de representar los distintos pasos que se siguen en un razonamiento basado en la revisión y fijación de creencias.

Las lógicas no-monótonas (en el sentido más amplio del término), teniendo en cuenta este tipo de restricciones han introducido la propiedad de la no-monotonía en las deducciones. Esta nos permite desechar una conclusión previa si se da el caso de que entra en nuestro sistema de conocimiento una información nueva conflictiva con la existente, en el sentido en que da lugar a la aparición de inconsistencias. Si bien esto funciona de acuerdo con nuestras restricciones, las cuestiones de contenido que delimitan la acotación de la información junto a sus mecanismos inferenciales no queda hoy por hoy clarificada; Reiter(1987), McCarthy(1980), Israel(1980), entre otros. No hay una fórmula de

circuscripción, u otro tipo de restricciones que nos permitan saber en cada momento qué información junto a qué inferencias deben ser utilizadas en el curso de un problema concreto, a excepción de los problemas en los que dispongamos de estrategias de control completas. Lo cual no indica que tales formalismos no deban ser desarrollados y utilizados, sino todo lo contrario.

En cuanto a la afirmación de que la lógica no es ni debería ser una teoría de la revisión/fijación de la creencia, sino una teoría de la implicación Israel(1980) - es decir, que la lógica está en relación con las consecuencias necesarias que se siguen de los hechos, mientras que el razonamiento y la epistemología deberían estar en relación con las reglas para derivar conclusiones "legítimas" a partir de los hechos- hay que ser muy cautelosos. La crítica de Israel, si estamos en lo cierto, se referiere a la imposibilidad de la lógica (por sí misma) para representar los aspectos intencionales presentes en un razonamiento, pero por ello no se induce que no puedan ser representados los procesos como tales, una vez que han sido introducidos los cambios oportunos en un modelo más amplio que mantuviera, sin embargo, al mismo lenguaje lógico base. En esta línea encontramos los trabajos de Wilensky(1983), Levesque(1984,87), Konolige(1984,91), Fagin y Halpern(1988), Pollock(1992) etc. Estos autores desarrollan la semántica formal y/o sintaxis de modo que tenga cabida la representación de las teorías de la creencia y de la intención. Una cuestión es representar la intención y otra bien distinta es dar cuenta del concepto de intencionalidad en un sentido filosófico tal y como lo hicieron Brentano o Husserl, pongamos por caso.

Por otra parte, es nuestra intuición la que nos indica que son precisamente los aspectos discursivos ligados al contenido, los que tienen la respuesta para este tipo de problemas intrínsecos a las teorías puramente formalistas cuando tratan de servir como modelos, en contextos que van más allá de lo puramente formal. El hecho de optar por un lenguaje u otro para la representación de las expresiones ya supone un cambio en cuanto a propiedades metalógicas, de expresividad, computacionales... Y esta elección parece responder más bien a motivaciones pragmáticas o de aplicabilidad que a cuestiones de índole lógica. Otro análisis que da luz al problema proviene de la 'diferenciación' entre razonamiento teórico y razonamiento práctico. Pensemos que la teoría de la racionalidad actual está

desarrollando mecanismos que permiten establecer una estrecha vinculación entre estos dos tipos de razonamiento. De hecho, no podemos diferenciar de forma tajante y definitiva entre ambos tipos de razonamiento. Usualmente se ha tendido a pensar que este tipo de razonamiento 'reevaluable' teórico podría quedar vinculado a (los estudios de) las distintas lógicas no monótonas o similares, mientras que la teoría del razonamiento reevaluable práctico estaría en conexión con las teorías de la revisión y fijación de creencias que incorporan mecanismos a modo de operadores formales para representar a las actitudes proposicionales. Si bien esto puede ayudar en cuanto a la comprensión intuitiva del proceso, una explicación más sutil del fenómeno pone de relevancia que ambos razonamientos son inseparables e interactúan de forma continuada. Y que es precisamente el razonamiento práctico el que controla el proceso y mediatiza al razonamiento teórico.

7. Conclusiones

Entre otras cosas, a lo largo de este artículo se ha ido mostrando que un enfoque exclusivamente sintáctico o semántico-formal en el tratamiento del razonamiento ordinario -tanto *modelos mentales* como *modelos formales*- no es suficiente desde el punto de vista de una Ciencia Cognitiva que pretenda crear modelos computacionales de los distintos procesos cognitivos con unas mínimas garantías de éxito. Los contenidos juegan un papel que no puede ser quizás recogido del todo por los mecanismos formales de inferencia sean de la naturaleza que sean. Con ello no queremos decir que no sea necesaria su aplicación, sino que junto a ellos, formando un "tandem", deben incorporarse los aspectos semánticos en la medida de sus posibilidades, dado que sólo de esta forma se alcanza el esclarecimiento conceptual necesario. Estos últimos requieren modos de representación más complejos que los ofrecidos hasta ahora por las distintas lógicas. Como consecuencia de ello situamos a las relaciones de contenido fuera de los mecanismos sintáctico-formales incidiendo en el hecho de que pueden influir en la elección de la información, en su acotación, así como en el establecimiento de criterios preferenciales cuando se trata de elegir entre varias reglas de inferencia, pongamos por caso. En este punto coincidimos con la mayoría de los psicólogos cognitivos, pero no en el hecho de que estas restricciones "semánticas" cierren el paso al estudio puramente lógico (competencial) de los procesos de razonamiento. Está claro que de los

mismos mecanismos sintácticos no podemos sin más establecer órdenes preferenciales sobre un conjunto de modelos minimales, quizás incompatibles entre sí respecto a una teoría, pero lo que sí permiten estudiar son las propiedades metalógicas que se seguirían manteniendose ciertas *condiciones de neutralidad*. Otra cuestión es que nos interese igualmente incorporar a nuestro modelo de razonamiento teórico (construido autónomamente) las características propias de las actitudes proposicionales, de modo que recubramos lo que hemos denominado teoría de la racionalidad general.

Por su parte, la representación es un tema con multitud de preguntas abiertas que esperan solución. Pensemos que si de un problema concreto se trata, entonces generalmente podemos representar determinada información almacenada en nuestra memoria, desechando el resto, asumiendo que la delimitación llevada a cabo es la adecuada con vistas a solventar el problema planteado. Es más, probablemente el problema nos traerá la cuestión enmarcada en el ámbito de un tipo de representación concreta; de alguna manera la respuesta podría llevarse a cabo dentro de los mismos límites de ese lenguaje elegido. Imaginemos, por ejemplo, cualquier problema aritmético básico; aquí las pautas a seguir vienen dadas en el propio lenguaje. No tendríamos más que utilizar las reglas aritméticas correspondientes para dar cuenta de cualquier problema, siempre, claro está, que mantuviéramos unas ciertas restricciones espacio-temporales.

El problema serio se plantea cuando queremos construir un lenguaje en el que representar problemas de carácter general, por ejemplo, para resolver el problema aquí tratado del razonamiento. Desde el punto de vista de la teoría del significado parece improbable que exista un medio representacional que nos permita dar cuenta del significado de un concepto dadas la insuficiencia de recursos; en relidad deberíamos enumerar un conjunto infinito de propiedades que sin lugar a dudas sobrepasarían nuestras capacidades computacionales. La idea es por tanto representar de forma abstracta y general, a modo de "modelo" o estereotipo que nos permita su utilización en situaciones diversas con un menor coste de recursos.

Si nos detenemos un momento en la propia idea de significado, veremos inmediatamente la imposibilidad de su definición. Si quiero dar cuenta del significado del sustantivo "juego" deberé enumerar todas sus propiedades

intrínsecas, lo que supone considerar un número ingente de variables que producen de inmediato la explosión combinatoria de posibilidades corvirtiendo el problema en intratable. Es decir, debería clarificarse su significado teniendo en cuenta lo que esto "significa" para cada uno de nosotros individualmente, notificando a su vez que su significado podría diferir en función del contexto, y además debería captar estos factores en cada una de las comunidades en las que se haga uso de la palabra "juego". Todo esto convierte en indefinible a su significado "global". Por ello en algún sentido debemos utilizar este concepto de modo que nos sea de utilidad en nuestro contexto social o familiar. La forma más plausible se nos antoja no muy lejana a la idea de estereotipo o algo similar. Es imposible definir un objeto en un sentido "sustancial". Sin embargo, podemos dar con una identificación del mismo que nos sea útil. ¿Por qué hablar de la idea de significado referida a un problema como el razonamiento ordinario?. Si hemos entendido correctamente lo aquí expuesto, no deberíamos tener dificultad a la hora de reconocer que la idea de representar los contenidos semánticos de las expresiones, aun siendo vital para la creación de un modelo cognitivo que pretenda mantener cierta plausibilidad desde el punto de vista de la Ciencia Cognitiva, no tiene por qué incidir sobre un tratamiento sintáctico-formal autónomo de lo que hemos denominado deducción, desarrollado por las distintas lógicas no-monótonas o similares. En resumidas cuentas, la cuestión básica subyacente no es la de establecer diferencias entre los modelos mentales y los modelos formales (es más, en muchos casos presentan una estructura parecida Shoham(1990)), sino más bien la de clarificar el papel que juegan las *actitudes proposicionales* cuando se trata de crear un modelo cognitivo completo del razonamiento. Sin duda es necesaria cierta motivación extra-lógica. Ahora bien, la motivación por sí misma no parece ser suficiente.

Bibliografía

R.J.BRACHMAN & H.J.LEVESQUE(1984), *A Fundamental tradeoff in Knowledge Representation.* **Procc. CSCSI-84,** London, Ontario; pp. 141-152.

C.CHERNIAK(1986), **Minimal Rationality**. Cambridge, Mass. : MIT Press.

N.CHOMSKY(1968), ***Language and Mind*** . New York: Harcourt, Brace & World.

P.R.COHEN & H.J.LEVESQUE(1987), *Persistence, Intention, and Commitment.* SRI International. Technical Note 415.

D.W.ETHERIGTON(1988), ***Reasoning with Incomplete Information.*** London : Pitman.

R.FAGIN and J.Y.HALPERN(1988), *Belief, Awareness, and Limited Reasoning,* **Artif. Intell. 34,** 39-76.

J.FODOR & PHYLYSHYN(1.988), ***Computación y Conocimiento. Hacia una Fundamentación de la Ciencia Cognitiva.*** Madrid.: Debate.

G.HAGERT and Y.WAERN(1986), *On Implicit Assumptions in Reasoning*, in : T.Myers, K.Brawn, and Mcgonigle (eds.). ***Reasoning and Discourse Processes***. London : Academic Press.

D.ISRAEL(1980), *What's Wrong with Non-Monotonic Logic*, in : **Proc. Amer. Assoc. for Artificial Intelligence-80**, pp. 99-101.

P.N.JOHNSON-LAIRD(1983), ***Mental Models : Towards a Cognitive Science of Language, Inference, and Consciousness***. Cambridge : Cambridge University Press.

P.N.JOHNSON-LAIRD(1986), *Reasoning without Logic,* in : T.Myers, K.Brawn, and McGonigle (eds.). ***Reasoning and Discourse Processes.*** London : Academic Press.

P.N.JOHSON-LAIRD(1988), ***The Computer and the Mind : An Introduction to Cognitive Science.*** London : Fontana.

W. and M.KNEALE(1961), ***El Desarrollo de la Lógica.*** Madrid : Tecnos.

K.KONOLIGE(1984), *Belief and Incompleteness*. Menlo Park, Ca. SRI International.

K.KONOLIGE(1991), *Intention, Commitment and Preference.* Menlo Park, Ca. SRI International.

J.LEE(1987), *Johnson-Laird's Models and Truth : A Discussion Paper.* Centre for Cognitive Science. University of Edinburgh.

H.J.LEVESQUE(1984), *A Logic of Implicit and Explicit Belief*, in: **Proceedings AAAI-84,** Austin, TX, pp. 198-202.

H.J.LEVESQUE(1988), *Logic and the Complexity of Reasoning*. **Philosophical Logic, 17,** 355-389.

C MALCOLM(1990), *Giving Ideas to Machines.* University of Edinburgh : AI Department.

J.McCARTHY(1980), *Circumscription - A form of Non-monotonic Reasoning.* **Artif.**

Intell. 13, 27-40.

D.McDERMOTT & J.DOYLE(1980), *Non-Monotonic Logic*. **Artif. Intell.13**, 41-72.

M.MINSKY(1981), *A Framework for Representing Knowledge*, in : J.Haugeland (ed.), ***Mind Design***. Cambridge, Mass. : MIT Press.

R.MONTAGUE(1974), ***Formal Philosophy***. New Haven : Yale University Press. (Trad. cast. *: **Ensayos de Filosofía Formal***. Madrid : Alianza

N.J.NILSSON(1991), *Logic and artificial intelligence*. **Artif. Intell. 47**; 31-56.

J.POLLOCK(1991), *A Theory of Defeasible Reasoning*. **International Journal of Intelligent Systems 6**, 33-54.

J.POLLOCK(1992), ***Cognitive Carpentry: a Blueprint for How to Build a Person*** (forthcoming).

H.PUTNAM(1988), ***Representación y Realidad***. Madrid : Gedisa.

R.REITER(1987), *A Logic for Default Reasoning*. **Artif. Intell. 13**, 41-72.

R.C.SCHANK(1981), ***Language and Memory***. New Haven : Yale University Press.

R.C.SCHANK(1982), *Dynamic Memory : A Theory of Reminding and Learning in Computers and People*, in : J.Haugeland (ed.) , **Mind Design**. Cambridge, Mass. :MIT Press.

Y.SHOHAM(1988), ***Reasoning about Change*** . Cambridge, Mass. :MIT Press.

Y.SHOHAM(1990), *Agent Oriented Programming*. **Report No. STAN-CS-90-1335**, October 1990. Dept. of Computer Science. Stanford University.

B.C.SMITH(1982), *" Reflection and Semantics in a Procedural Language"*. PH.D.Thesis. M.I.T., Cambridge, Mass. Reprinted in R.J. Brachman and H.J. Levesque (Eds.) (1985), **Readings in Knowledge Representation**. Los Altos, Ca.: Morgan Kaufman.

K.STENNING(1990), *Modelling Memory for Models*. Human Communication Research Centre. University of Edinbugh.

R.WILENSKY(1983), ***Planning and Understanding : A Computational Approach to Human Reasoning***. Reading, Mass. : Addison-Wesley.