EPISTEMOLOGÍA E HISTORIA DE LA CIENCIA

SELECCIÓN DE TRABAJOS DE LAS VII JORNADAS

1997

Patricia Morey
José Ahumada
Editores

ÁREA LOGICO-EPISTEMOLÓGICA DE LA ESCUELA DE FILOSOFÍA
CENTRO DE INVESTIGACIONES DE LA FACULTAD DE FILOSOFÍA Y HUMANIDADES
UNIVERSIDAD NACIONAL DE CÓRDOBA

Claudio Delrieux
U. N. del Sur
e-mail:claudio@acm.org

REPRESENTACIÓN DE CONOCIMIENTO Y RAZONAMIENTO EN LOS PROGRAMAS DE INVESTIGACIÓN DE LAKATOS

Palabras Clave: Inteligencia Artificial - Sistemas Basados en Conocimiento - Teoría de la Ciencia - Programas de Investigación Científica --- Razonamiento no Monotónico

Resumen

La implementación computacional de sistemas basados en conocimiento requiere la solución a determinados problemas de representación y razonamiento de larga data, específicamente los problemas relacionados con el conocimiento no deductivamente válido. Uno de los ejemplos prácticos más complejos es el razonamiento científico, donde se presenta una diversidad de elementos constitutivos y de contextos de funcionamiento que determina una elaborada estructura de representación de conocimiento y razonamiento. En este trabajo presentamos un análisis de los *programas de investigación* de Lakatos elaborado desde el punto de vista de la Inteligencia Artificial. Se consideran los diversos elementos que constituyen los programas de investigación a partir de una estructuración epistémica, y se estudian los diversos procedimientos de inferencia como casos especiales de razonamiento no monotónico.

1 Introducción

La Inteligencia Artificial (IA) propone nuevas pautas y puntos de vista con respecto a los problemas epistemológicos tradicionales. En efecto, la implementación de sistemas basados en conocimiento requiere una solución computacionalmente adecuada para los problemas de representación de conocimiento y razonamiento (KR&R). Ésto obliga, entre otras cosas, a dar definiciones de *conocimiento* que sean operacionalmente útiles, además de ser formalmente correctas (Israel83). La forma usual y menos

problemática de representar conocimiento analítico (universalmente válido), consiste en utilizar fragmentos decidibles de un lenguaje lógico de primer orden clausurado bajo deducción (McCarthy89). Otros *tipos* de conocimiento y sus modos asociados de razonamiento, sin embargo, quedan fuera de las posibilidades de esta representación. Pero estos tipos de conocimiento son indispensables para la mayoría de los problemas que enfrenta un sistema basado en conocimiento. La causa de esta inadecuada exclusión puede rastrearse históricamente hacia los orígenes de la formalización matemática de la lógica, específicamente a los trabajos de Frege, Russell y Tarski. La misma deja de lado toda una tradición filosófica de valorar el conocimiento en función de su estructura interna, de su origen o de su justificación.

Un caso paradigmático de la necesidad de incluir estos tipos de conocimiento lo podemos ver en la teoría de la ciencia. Las teorías científicas, a diferencia de las teorías lógicas, incorporan un conjunto heterogeneo de conocimiento, el cual está jerárquicamente organizado en función del contexto en el que se desenvuelve. Esta organización refleja una dimensión estratégica, contemplando el posible beneficio del *uso* de dicho conocimiento por medio de determinados patrones de inferencia englobados en lo que comunmente se denomina *método*. En el quehacer científico cotidiano, el énfasis no está puesto en la sintaxis ni en la semántica, sino en los aspectos pragmáticos. No es esperable que una teoría científica durante su desarrollo obedezca ciertos principios metateóricos como la consistencia (dado que en principio debe aceptarse toda información) o la completitud. La inferencia deductiva, si bien es esencial, no es excluyente como en las teorías lógicas, sino que cumple un rol más bien rutinario, siendo otro tipo de procedimientos de inferencia (por ejemplo abducción, inducción o razonamiento hipotético) los que cumplen un papel destacado. Estos procedimientos normalmente son no correctos (unsound), pero sí son de gran valor heurístico.

La formalización de estos aspectos de la teoría de la ciencia enfrenta un conjunto de dificultades que son consecuencia de la estructura del conocimiento científico. Todas las características mencionadas determinan que desde el punto de vista lógico, la representación adecuada de una teoría científica sea de gran dificultad. Sin embargo, el objetivo tiene sus beneficios, porque una adecuada formalización provee un marco sistemático e inambiguo para describir las estructuras y los mecanismos involucrados, y es una demostración de que los mismos son coherentes y operacionales. Además, una formalización que trate de mantenerse dentro de las estructuras y procedimientos constructivos y decidibles, es en principio computacionalmente tratable, por lo que su implementación puede arrojar resultados novedosos.

En este trabajo desarrollamos como ejemplo ilustrativo una posible formalización de una teoría de la ciencia, inspirada en los *programas de investigación científica* de Lakatos. La presentación original de los *programas* (Lakatos70, 78) está dada en términos muy informales, por lo que aquí se propone la introducción de determinados conceptos basados en los grandes puntos de contacto entre el razonamiento científico y el razonamiento no monotónico. La formalización presentada en este trabajo comienza por caracterizar distintos *tipos* de conocimiento que se utilizan en una teoría científica en desarrollo. Ésto permite considerar al comportamiento de un *programa* como un caso particular de razonamiento no monotónico con importancia epistémica. Nuestro desarrollo

incorpora algunos resultados relevantes de la investigación en razonamiento no monotónico, donde se utiliza conocimiento general tentativo en forma de condicionales o *reglas derrotables* (Reiter80, Poole85, Loui87). Estos condicionales derrotables permiten representar en la teoría las *hipótesis contrastadoras* y *generalizaciones accidentales* (Hempel65). Además se incorpora el manejo de la evidencia como conocimiento particular tentativo para representar los postulados hipotéticos auxiliares. Puede establecerse un ranking de preferencia entre las distintas piezas de conocimiento, reflejando la distinta importancia epistémica que el programa de investigacion asigna a cada una en función de una determinada estrategia. El razonamiento con dicho tipo de conocimiento es similar al razonamiento plausible presentado por varios autores, especialmente Rescher (Rescher74, 76). Este razonamiento se basa en un tipo de conocimiento que permite representar resultados experimentales, postulados hipotéticos y toda otra información particular tentativa como por ejemplo lo que normalmente constituye el *cinturon protector* de una teoria cientifica (Lakatos70). Esta formalización puede considerarse como una extensión a trabajos presentados anteriormente (Delrieux95a, 95b).

La estructura del trabajo es la siguiente. En la próxima sección introduciremos algunos elementos de la teoría de la ciencia desde el punto de vista de la Inteligencia Artificial. En la sección 3 se presentan algunos conceptos fundamentales de la presentación de Lakatos de los programas de investigación. En la sección 4 se describe formalmente la representación de los distintos elementos de un programa de investigación definiéndose los diversos tipos de conocimiento que concurren en una teoría, así como su estructuración por medio de una relación de preferencia. Se discuten también los patrones de inferencia asociados al contexto de explicación y de predicción para, en la sección 5, proponer una posible representación del comportamiento de los programas frente a contrastaciones negativas, la comparación de teorías, y la generación de nuevo conocimiento. Por último, en la sección 6 se presentan las conclusiones y las líneas de trabajo futuro.

2 La Teoría de la Ciencia desde el punto de vista de la IA

El propósito de incluir conocimiento en los sistemas de IA es permitir una *representación* más o menos precisa y completa, dentro del sistema, de un determinado dominio. Una enunciación explícita de este acercamiento fue propuesta en IA por primera vez por John McCarthy (McCarthy69, 77), quien fue sin duda una de las influencias más importantes, al dar forma a toda el área de representación de conocimiento y razonamiento (KR&R). Según McCarthy, una entidad es *inteligente* si tiene un modelo adecuado del mundo, incluyendo, además de lo fáctico, los aspectos formales y otros procesos intelectuales, como por ejemplo sus propias intenciones. Para dicha representación, la herramienta por excelencia es la lógica. El lenguaje de la lógica permite representar enunciados con función informativa. Dentro de este contexto, la lógica en un sentido amplio puede verse como el estudio de los patrones de inferencia del sentido común.

En el área de KR&R se utilizó la lógica matemática (Frege-Russell-Tarski), ignorando toda una tradición filosófica en el estudio del conocimiento, especialmente en la consideración de diversos *tipos* de conocimiento, según sea su estructura, origen, justificación, etc. Sin embargo, estos tipos de conocimiento son de una importancia central

para los problemas que enfrenta el área de KR&R. Ésto sucede por ejemplo en la implementación computacional del una teoría de la ciencia. Las teorías científicas son conjuntos heterogeneos de conocimiento, los cuales están estructurados de acuerdo a una jerarquía, y cuyo objetivo consiste en sistematizar, predecir o explicar un determinado conjunto de fenómenos existentes en la realidad. Dentro del procedimiento científico, especialmente dentro de las ciencias experimentales, se estableció un conjunto de procedimientos que permiten llevar adelante en forma adecuada la explicación y predicción de fenómenos, la generación de teorías, y la corroboración y refutación de las mismas. La formalización de estos procedimientos constituye el objeto de estudio de la Teoría de la Ciencia.

El conocimiento científico se ordena y configura en estructuras complejas. Las unidades de organización y análisis más destacadas son las *teorías*. Las teorías científicas tienen la función de establecer conexiones sistemáticas dentro de un aspecto de la realidad. De ese modo es posible la inferencia de determinados hechos a partir de otros. Es importante destacar la gran similitud y al mismo tiempo gran diferencia entre teorías científicas y teorías lógicas. Si el *tipo* de conocimiento que constituye una teoría científica fuese conocimiento verdadero, entonces no habría diferencia entre ambos tipos de teorías. Sin embargo, las teorías científicas involucran tipos de conocimiento cuya justificación es problemática, y por lo tanto no tienen el *status* de ser analíticas o deductivamente válidas. Ésto establece asimetrías en el comportamiento de los patrones de inferencia asociados a estas teorías.

Podemos describir por lo menos tres dominios o niveles de enunciados dentro de una teoría (Gianella95, Klimovsky95). El primer nivel *N1* es un conjunto de enunciados empíricos particulares que representa los diversos estados de cosas posibles en dicho dominio. Un enunciado cualquiera *p(a)* se interpreta como *"es un hecho empiricamente observable que el objeto (entidad, fenómeno) a tiene la propiedad (característica, circunstancia) p"*. Normalmente los enunciados de este nivel asumen la forma de literales de base, donde tanto los predicados (que representan propiedades, características, etc.) y los términos (que representan objetos, entidades, etc.) son observables. El segundo nivel *N2* está constituido por generalizaciones empíricas o accidentales. El objetivo del conocimiento en este nivel es representar de una manera regular y económica las clasificaciones o correlaciones que se han podido observar en conjuntos de enunciados del nivel anterior. Un enunciado de este nivel adopta la forma de una *ley* (lawlike statement) universal, existencial o probabilística, pero referida a términos y relaciones observables, por ejemplo *"Algunos (todos los) objetos (entidades, fenómenos) que tienen la propiedad (caracterisitica, circunstancia) observable p, normalmente tienen la propiedad q"*. El tercer nivel *N3* contempla los enunciados *teóricos*, es decir, representa el conocimiento de aquellos elementos de la teoría que no son estrictamente observables. Estos enunciados teóricos son denominados tambien principios internos. Este nivel es el más importante de una teoría, pues es el que le confiere su identidad como tal, y permite dar cuenta en profundidad de lo que se conoce en los niveles anteriores. Los enunciados en este nivel normalmente son leyes universalmente cuantificadas.

Las teorías forman parte de una disciplina, pero no las sistematizan exhaustivamente. Un posible propósito para las teorías es buscar el menor conjunto de

conocimiento que produzca un *cubrimiento* del conjunto de evidencia E que se pretende sistematizar. Este cubrimiento se produce a través de un conjunto de procedimientos de inferencia. Las primeras descripciones (por ejemplo las del Círculo de Viena) se basaron fundamentalmente en justificaciones inductivas. El esquema subyacente consistía en mostrar que las leyes científicas se infieren inductivamente a partir de la evidencia. Este acercamiento fue encontrando dificultades insalvables. Hempel (Hempel48) fue el primero en proponer que la evidencia debe inferirse de las leyes, y no a la inversa. Según sea que la inferencia se haya realizado antes o después que los hechos deducidos se hayan comprobado, la misma se denomina *predicción* o *explicación*. La lógica de la predicción y de la explicación proceden según un mismo esquema $L|\text{-}e$, donde L, el *explanans*, es un conjunto de leyes, y e, el *explanandum*, es el fenómeno o hecho a explicar. La única diferencia constituye el *contexto* dentro del cual se utiliza el esquema, el cual, siguiendo a Reichenbach, es denominado *contexto de descubrimiento* y *contexto de explicación*, respectivamente. Es conveniente notar que en este esquema, el *explanandum* pertenece al primer dominio o nivel ($e \in N1$), mientras que el *explanans* pertenece a los otros dos ($L \in (N2 \cup N3)$).

La sistematización por medio de este esquema se denominó *paradigma hipotético-deductivo*, dado que el *explanans* constituye una pieza de conocimiento hipotético, del cual se debe deducir la evidencia. Uno de los criterios pragmáticos expresados en al paradigma hipotético-deductivo consiste en justificar el fracaso de una determinada ley frente a una contrastación dada, no por ser falsa dicha ley, sino por ser inaplicable para ese caso particular. Formalmente ésto se consigue debilitando el *explanans* con una sentencia particular $c \in N1$, que hace referencia a ciertas condiciones particulares relevantes para la evidencia e a explicar en esta contrastación. Formalmente el esquema deviene en una *implicación contrastadora* $L|\text{-}(c \rightarrow e)$. El conjunto C de condiciones particulares a los que apela una teoría para efectuar un cubrimiento constituye el conjunto de hipótesis particulares de la teoría. Entonces existen dos subconjuntos del nivel $N1$ de enunciados particulares, la evidencia-dato y la evidencia-resultado. El cubrimiento por leyes, desde este punto de vista, significa que las leyes L son tales que cada vez que se dan las condiciones particulares $C \subseteq N1$, entonces se predicen o explican los estados de cosas $E \subseteq N1$.

El procedimiento de inferir el *explanans* no puede ser deductivo, es decir, L nunca puede ser *verdadera*. Una conclusión, señalada sistemáticamente por Popper (Popper59) es que las teorías científicas no se *verifican* sino que se *refutan*. Dicho de otra forma, no existe evidencia posible que garantice la verdad lógica de una teoría, pero una sola predicción o explicación incorrecta -aunque sea frente a una cantidad enorme de casos correctos- sirve para mostrar que una teoría es falsa. Este comportamiento demuestra que el esquema hipotético-deductivo es pragmáticamente poco adecuado. Como veremos más adelante, mientras una teoría produzca resultados positivos no será completamente abandonada. Este hecho, observado por Lakatos (Lakatos70), fue el inspirador de su reconstrucción de la dinámica de las teorías científicas, denominadas por él *programas de investigación*, de la que nos ocuparemos en detalle en la próxima sección.

3 Los programas de investigación científica

Un punto de vista importante en Lakatos es tener en cuenta el aspecto dependiente del contexto histórico que tiene la dinámica de una teoría científica. Por dicha razon, se separa *historia interna* de la *historia externa* en la evolución de las teorías. La historia externa es el registro histórico o periodístico de la secuencia de hechos relevantes en el desarrollo de una teoría desde su origen hasta su abandono total. La historia interna, en cambio, registra los elementos de juicio desde el punto de vista de la comunidad que participa de dicho desarrollo, y es por lo tanto una reconstrucción racional de la historia externa. Este aspecto histórico determina que el *método* científico, en el quehacer cotidiano, no sea único, es decir, distintas comunidades adoptan distintas prácticas o siguen distintas estrategias, hasta que eventualmente se agotan. Por lo tanto, en una misma ciencia pueden coexistir diferentes teorías para explicar un mismo fenómeno, cada una propugnada por una parte de la comunidad científica que adhiere a un determinado aspecto metodológico o defiende determinadas piezas de conocimiento en detrimento de otras. Cada una de estas teorías, junto con su metodología subyacente, es un *programa de investigación* que compite con los demás. Normalmente estos programas no se demuestran acertados o equivocados en el tiempo, sino que se ven superados por nuevos descubrimientos. Una teoría no puede ser absolutamente verdadera como lo es un enunciado analítico. Es más, una teoría puede ser falsa pero tener consecuencias verdaderas y operacionales. Por esta razón los programas de investigación permanecen abiertos y sujetos al cambio y la evolución. Este proceso, sumado a la competencia por la supervivencia de los distintos programas (posiblemente determinada por las condiciones económicas y culturales), es totalmente análogo a la evolución natural. En ésto radica su eficacia. Un sistema selectivo discrimina sus distintos elementos en función de propiedades o características ventajosas. En el caso de la investigación, se seleccionan hipótesis, conjeturas, y hasta teorías completas. La presión competitiva, el mecanismo de selección, proviene de la realimentación crítica. Ésto determina que un programa pueda sobrevivir sin que por ello sea verdadero.

Los programas de investigación son estructuras que incluyen a las teorías científicas, integrándolas con un conjunto de procedimientos de inferencia y que poseen un conjunto periférico de hipótesis auxiliares. El *núcleo* de un programa es un conjunto de conocimiento que se considera central, y que define la teoría como tal. Este núcleo es el conjunto de conocimientos (leyes, generalizaciones o postulados) que determina la identidad de la teoría y por consiguiente del programa mismo. El núcleo, por lo tanto, se considera definitivo, y el resto de la estructura del programa opera de modo tal de protegerlo de la refutación. Esta protección consiste básicamente en implementar un *cinturón protector* de hipótesis auxiliares, que impiden que el núcleo sea refutado. De esa forma, existen dos procedimientos heurísticos para confrontar a la teoría con la evidencia de un resultado experimental *e*. Si el resultado *e* no es correctamente predicho o explicado por la teoría, entonces se aplica la heurística negativa, que consiste en encontrar una hipótesis *c* particular al caso *e* tal que de la teoría aumentada con *c* se siga *e*. Si dicha hipótesis no es compatible con el resto de la teoría, entonces algo en la misma deberá corregirse. Si recordamos la observación efectuada en la sección anterior con respecto a la evidencia dato y evidencia resultado, una estrategia posible para esta heurística es negar que determinada

evidencia dato sea adecuada. Ésto concuerda, como vimos, con el procedimiento Hempeliano de la implicación contrastadora, y en ese sentido puede afirmarse que el conjunto C de hipótesis auxiliares constituye un cinturón protector. Mientras no ocurran refutaciones, entonces se aplica la heurística positiva, en la cual se busca una sistematización del cinturón protector (como consecuencia del núcleo duro o por medio de nuevas leyes).

La importancia de todos estos elementos en la dinámica de un programa de investigación puede determinarse solamente dentro del marco selectivo expresado más arriba. Si existen dos o más programas en competencia, probablemente el más exitoso sea aquel cuyo cinturón protector sea menor, aunque las leyes que conforman su núcleo no sean aún totalmente aceptadas. Un ejemplo histórico bien conocido es la crisis de la mecánica newtoniana ocurrida a principios de siglo. Los experimentos de Michelson-Morley, la mecánica subatómica y ciertos fenómenos astronómicos como la precesión de los equinoccios de Mercurio, contradecían las leyes de Newton, y para ser explicados requerían más y más hipótesis *ad-hoc*. La teoría de la Relatividad de Einstein surge como competidora de la mecánica newtoniana, con una ley general que explicaba todos esos fenómenos: la velocidad de la luz es un invariante. Esta teoría fue rápidamente adoptada, pese a que pasaron varios años hasta que se pudo constatar el único fenómeno experimental predicho exclusivamente por la Relatividad: la curvatura del espacio por acción de la gravitación.

Uno de los aspectos que Lakatos no cubrió, tal vez por su temprana muerte, fue una formalización adecuada de los elementos que constituyen un programa de investigación. Dicha formalización constituye uno de los objetivos de este trabajo. Más aún, la presentación aquí planteada es puesta en concordancia con muchos aspectos del razonamiento no monotónico. Esto permitirá, por fin, esbozar los lineamientos generales de la implementación computacional de muchos aspectos de una teoría de la ciencia. La importancia de dicha formalización proviene de que provee un marco sistemático e inambiguo para describir las estructuras y los mecanismos involucrados (que como vimos son bastante complejos), y es una demostración indirecta pero eficaz de que los mismos son coherentes y operacionales.

4 La formalización de un programa de investigación

Nuestra descripción formal establece que un programa de investigación está conformado por una *estructura epistémica* compuesta por subconjuntos de los distintos *tipos* de conocimiento disponibles, una *preferencia* que determina la diferente importancia que tiene cada pieza de conocimiento dentro de la teoría, y por un por un conjunto de *procedimientos de inferencia* que son utilizados en función del contexto en el cual trabaja el programa. El lenguaje de la lógica clásica no es lo suficientemente expresivo como para poder representar formalmente los elementos que constituyen un programa de investigación. Por dicha razón primero veremos una serie de puntos de contacto entre el razonamiento científico y el razonamiento no monotónico. El razonamiento científico es no monotónico, porque muy raramente los resultados que produce son firmes.

El conocimiento del sistema se representa en un un lenguaje de primer orden en el que es posible efectuar inferencias deductivas. Utilizaremos la convención de nombrar con letras minúsculas italizadas (por ejemplo *a* o *b(X)*) para referirnos a proposiciones o sentencias arbitrarias en dicho lenguaje. Utilizaremos mayúsculas italizadas en negrita (por ejemplo ***K***) para referirnos a conjuntos de sentencias generales (nonground), y con mayúsculas italizadas (por ejemplo *C*) para referirnos a conjuntos de sentencias particulares (ground). Dicho lenguaje es extendido para permitir la representación de piezas tentativas de conocimiento. Las sentencias que representan conocimiento tentativo general asumen la forma de condicionales o implicaciones *prima facie*. Por ejemplo, la expresión $a(X) \triangleright b(X)$ expresa que *"La disposición de aceptar a(X) es una razón para aceptar tentativamente b(X)"*. Este tipo de condicionales expresan una verdad tentativa referida a propiedades, por lo que puede considerarse una forma de conocimiento modal *de re* (Chisholm77). Las sentencias que representan conocimiento tentativo particular, que asume la forma de evidencia tentativa, se representan como literales indexados l_i, que expresan la disposición a considerar el conocimiento particular *l*, al provenir éste de un critero tentativo *i* (información exacta o inexacta, conjetura, criterio estadístico, etc.). Estas sentencias proveen información referida a estados de cosas, por lo que pueden considerarse una forma de conocimiento modal *de dicto*.

4.1 Teoría = conjuntos de conocimiento + preferencia epistémica

En nuestra definición, una *teoría científica* ***T*** está constituída por la unión de enunciados pertenecientes a los siguientes conjuntos de conocimiento: ***K***, *conocimiento lógico-matemático* deductivamente válido; ***P***, los *principios internos* de la ciencia en cuestión; ***H***, las *hipótesis explicativas* que se derivan de ***P*** y forman parte del *núcleo* de la teoría; ***G***, las *generalizaciones accidentales* que surgen como abstracción de un conjunto razonablemente grande de casos particulares, y que tambien conforman el núcleo de la teoría; *E*, la *evidencia* es el conjunto de datos experimentales que la teoría utiliza; y *C*, las *hipótesis auxiliares* particulares, utilizadas junto con las hipótesis explicativas para predecir o explicar piezas de evidencia. Los primeros cuatro conjuntos representan conocimiento general (nonground). ***K*** y ***P*** se refieren a entidades no observables, mientras que ***H*** y ***G*** normalmente tienen contenido empírico. ***K*** puede interpretarse como la conjunción de todas las sentencias que expresan el conocimiento lógico-matemático. Las leyes de la lógica quedan incorporadas de modo tal que el lenguaje sea deductivamente cerrado.

P expresa los principios internos de la disciplina en cuestión, y por lo tanto incluye todas las definiciones referidas a las entidades no observables de una ciencia, de las cuales todas las demás "leyes" son consecuencia. Por ejemplo, las ecuaciones de Maxwell del electromagnetismo son descripciones de la variación espaciotemporal de los campos asociados a una onda electromagnética. Al ser definiciones, normalmente expresan un conocimiento sintético acerca de determinados aspectos de la realidad, es decir, no se siguen deductivamente de ***K***. Por lo tanto, un programa de investigación solo puede tomar la decisión estratégica de aceptar o rechazar piezas de este tipo de conocimiento. No puede deducirlas a partir de otro conocimiento. ***H*** representa el conjunto de hipótesis explicativas

o "leyes", las cuales se siguen deductivamente a partir de K y P pero tienen un valor pragmático más adecuado al referirse a entidades observables. Por ejemplo, las leyes de la óptica geométrica (por ejemplo las de Snell) y las leyes de la electrotecnia (por ejemplo las de Ohm) se siguen lógico-matemáticamente de las ecuaciones de Maxwell. Sin embargo, ningún óptico o electrotécnico en su sano juicio utiliza las ecuaciones de Maxwell para resolver problemas que normalmente ocurren en la práctica.

G expresa el conjunto de generalizaciones accidentales que surgen de un proceso de abstracción a partir de la observación de un conjunto de casos particulares, por ejemplo que las resistencias eléctricas elevan su temperatura al disipar grandes potencias. Por lo tanto, si bien expresan un conocimiento general, están sujetas a excepciones, y luego no es adecuado representarlas dentro del lenguaje. Para su representación, entonces, utilizaremos las sentencias condicionales derrotables. Esto permite que puedan actuar como leyes *prima facie*. E, la *evidencia* es el conjunto de datos experimentales que la teoría utiliza como conocimiento contingente (literales de base) para poder derivar conclusiones referidas al dominio de las entidades observables. Por último, C expresa un conjunto de hipótesis auxiliares que se aplican a determinados casos particulares, para evitar el fracaso en predecir o explicar adecuadamente elementos del dominio de las entidades observables por parte de las hipótesis explicativas. Normalmente también son literales de base.

Definición 4.1: Dado un contexto K, P (el conocimiento lógico-matemático y los principios internos), una **Estructura Epistémica** $\mathbf{E}_{K,P}$ es una estructura de conocimiento $\mathbf{E}_{K,P} \subseteq \langle H,G,E,C\rangle$, donde H es un conjunto consistente de conocimiento intensional tal que $P \vdash H$, G es un conjunto finito de condicionales de la forma $a(X) \triangleright b(X)$, E es un conjunto conocimiento particular y C es un conjunto de hipótesis auxiliares representadas como conocimiento tentativo de la forma l_i, donde l son literales de base e i corresponde a un criterio de aceptación. Cuando el contexto quede claramente definido, nos referiremos a una estructura epistémica simplemente como $\mathbf{E}$. □

A diferencia de una teoría lógica, en una teoría científica existe necesariamente una estructuración jerárquica del conocimiento. Como es posible entrever, uno de los elementos esenciales en nuestra formalización de los programas de investigación consiste en representar esta relación de preferencia epistémica dentro de los elementos de conocimiento en una teoría. Es decir, el programa de investigación posee un criterio de comparación $\prec$ que le permite decidir si una pieza de conocimiento es preferible a otra por su importancia epistémica. Los únicos conjuntos de conocimiento firme, es decir, aquellas piezas de conocimiento que no pueden ser en principio cuestionadas, son K y E.

Definición 4.2: Dada una estructura epistémica $\mathbf{E}_{K,P}$, una **Teoría** T es un par $T=\langle \mathbf{E}_{K,P}, \prec\rangle$, donde $\prec$ es un orden parcial sobre los enunciados de T, llamado relación de **Preferencia Epistémica**. T contiene un elemento T_T tal que $\forall\alpha \in T.\alpha \prec T_T$ (por ejemplo, el conocimiento analítico o la evidencia debe obedecer esta condición), y un elemento $T_\perp$ tal que $\forall\beta \in T.T_\perp \prec \beta$. De esa manera T queda reticulada bajo $\prec$. □

Es importante destacar algunos puntos de estas definiciónes. Cada teoría selecciona enunciados del conjunto total de conocimiento $\langle H,G,E,C\rangle$ que es en principio definible dentro de una disciplina. Estos enunciados, como vimos, no tienen el requisito de ser consistentes entre sí, sino que cada uno de ellos debe ser simplemente consitente con

$K \cup E$. A este subconjunto de enunciados se le asigna *arbitrariamente* un orden parcial $\prec$ que lo estructura. El ordenamiento dependerá estratégicamente del contexto dentro del cual funcione la teoría. De esa manera, en una disciplina pueden coexistir varias teorías, sustentada cada una por una estructura epistémica distinta que la justifica. En el resto de este trabajo presentaremos una caracterización formal de los problemas más importantes que se plantean en el razonamiento científico: cuáles son las conclusiones (predicciones o explicaciones) que justifica una teoría, cómo se comparan teorías entre sí, y cómo es posible justificar la generación de nuevo conocimiento.

4.2 Programa de investigación = teoría + procedimientos de inferencia

El segundo paso en nuestra formalización consiste en definir a un programa como una teoría que progresa en función de determinados procedimientos de inferencia. Uno de los aspectos más importantes consiste en determinar cuál es el conjunto de conclusiones que se justifican a partir de una teoría ***T***, tanto en el contexto de predicción como en el de explicación. Otros procedimientos de inferencia, relacionados con las contrastaiones negativas, la comparación de teorías o la justificación de nuevo conocimiento en función de una "lógica" del descubrimiento, serán presentadas en la sección siguiente. Algunas de las ideas presentadas en esta subsección se basan en el sistema ***P*** de razonamiento plausible (Simari94, Delrieux95a). Dado que las teorías no son necesariamente consistentes, la idea esencial es que las conclusiones de una teoría son la consecuencia deductiva de los conjuntos máximamente consistentes de la misma.

Definición 4.3: Dada una teoría $T=< E_{K,P}, \prec >$, y un subconjunto T_1 de **E**, la **Importancia Epistémica** de T_1 se define como el conjunto $\left\{\alpha \in T_1 \middle| \nexists \beta \in T_1 . \beta \prec \alpha\right\}$ de cotas inferiores de T_1 bajo $\prec$. Dados dos subconjuntos de una teoría T_1 y T_2, diremos que T_1 es **epistémicamente más importante** que T_2 (denotado como $T_1 \prec T_2$) si y sólo si cada enunciado en T_1 es al menos tan importante en T como cada enunciado en T_2, pero existe por lo menos un enunciado en T_1 que es estrictamente más importante que cada enunciado en T_2. □

Dada una teoría ***T***, cuál es el subconjunto consistente de enunciados de ***T*** de mayor importancia epistémica? La solución aquí propuesta consiste en considerar la intersección de todos los conjuntos generados bajo distintas extensiones lineales de $\prec$.

Definición 4.4: Dada una teoría $T=<E, \prec>$, una **Extensión Lineal** e de $\prec$ es una relación que contiene a $\prec$ y que induce un orden lineal en **E**. □

Ejemplo 4.1: Supongamos que tenemos los enunciados **E** = {a, b, c} y que la relación de preferencia en E establece que {b $\prec$ a, c $\prec$ a}. Entonces tenemos dos extensiones lineales posibles para $\prec$, una en la cual {c $\prec$ b}, y otra en la cual {b $\prec$ c}.□

Definición 4.5: Dada una teoría $T=< E_{K,P}, \prec >$ y una extensión lineal e de $\prec$, un **Subconjunto Máximamente Consistente (SMC)** de T (con respecto al contexto K, P) es un conjunto E^e que satisface[1]:

- $E^e \subseteq E$ (es un subconjunto de la estructura epistémica),
- $E^e \cup K \cup E \not\vdash \bot$ (es consistente con el conocimiento matemático y la evidencia),
- $\forall \alpha \in E^e . \forall \beta \in (E / E^e). \beta \prec \alpha$ (los enunciados en E^e son los más importantes), y
- $\forall E'. E^e \subset E' \subseteq E, (E' \cup K \cup E) \vdash \wedge$ (es maximal).

El **conjunto de Conclusiones** (predicciones o explicaciones), entonces, es la clausura de la intersección de todos los SMC inducidos bajo toda extensión lineal de $\prec$. □

Es importante mencionar que existe procedimiento efectivo de prueba para determinar si una sentencia dada está en el conjunto de conclusiones de una teoría.

Definición 4.6: Dada una teoría $T=<E, \prec>$ y una consulta q tal que ni $K \cup E \vdash q$ ni $K \cup E \vdash \neg q$. Entonces definimos:

- **Fundamento:** q **está fundado** si existe un conjunto de fundamento $E_f \subseteq E$, tal que $E_f \cup K \cup E \vdash q$.
- **Duda:** q **está en duda** si existe un conjunto de duda $E_d \subseteq E$, tal que $E_d \cup K \cup E \vdash \neg q$.
- **Aceptación:** q **es aceptada** si está fundado y no está en duda, o bien si $E_d \prec E_f$, es decir, la importancia epistémica de su conjunto de fundamento es mayor que la de su conjuunto de duda.
- **Rechazo:** q **es rechazada** si está en duda y no está fundado, o bien si $E_f \prec E_d$, es decir, la importancia epistémica de su conjunto de duda es mayor que la de su conjunto de fundamento. □

Esta definición es computacionalmente tratable, ya que incurre en una doble recursión, y las invocaciones a demostrabilidad, al realizarse por encadenamiento hacia atrás, se computan con técnicas estándar de programación en lógica. El siguiente teorema muestra que el procedimiento de aceptación descripto es correcto y completo con respecto a la definición 4.5 de conclusiones de una teoría.

Teorema 1: Dada una teoría $T = <E, \prec>$, una consulta q es aceptada con fundamento E_f (con $E_f \subseteq E$) si y solo si q pertenece a la intersección de todos los subconjuntos máximamente consistentes de T bajo distintas extensiones lineales de $\prec$. □

Los detalles de la demostración pueden consultarse en (Delrieux95a). Un ejemplo del comportamiento de una teoría (desarrollado también en Delrieux95b) es el siguiente:

Ejemplo 4.2: Nuestro conocimiento acerca de gravitación se reduce a:

$P_1: \forall X,Y. (o(X) \wedge o(Y)) \Rightarrow a(X,Y)$ (Existe una fuerza que atrae a los objetos (masivos) entre sí.)

---

[1] Abusando de la notación, utilizaremos subconjuntos de la estructura epistémica como parte del antecedente del operador de consecuencia deductiva clásico, para expresar relaciones de consecuencia que resultarían si los enuciados de dichos conjuntos subconjuntos fueran utilizados por las reglas de inferencia de dicha relación de consecuencia, particularmente los miembros de G como implicaciones materiales y los miembros de C como literales.

H_1: $\forall X.\ o(X) \Rightarrow a(tierra,X)$ (Los objetos son atraídos hacia la tierra.)
e_1: $a(tierra,p)$ (Esta piedra es atraída hacia la tierra.)
e_2: $\neg a(tierra,g)$ (Este globo de hidrógeno no es atraído hacia la tierra.)

La primer sentencia es un principio, la segunda es una hipótesis explicativa, y las dos últimas son evidencia. La hipótesis explicativa se deduce como caso particular del principio, y es por lo tanto utilizada por su mayor valor pragmático. Sin embargo, si bien permite explicar (o predecir) la tercer sentencia, fracasa con la cuarta. Es decir, tenemos $\boldsymbol{P}_1 \vdash \boldsymbol{H}_1, \boldsymbol{H}_1 \vdash e_1$, pero $\boldsymbol{H}_1 \not\vdash e_2$. Ésto, sin embargo, no lleva a abandonar la teoría, sino a buscar razones por las cuales fracasa en la explicación de este caso. Es decir, la teoría se modifica de modo tal que la justificación tome la forma $\boldsymbol{H}_1, C \vdash E$, donde E cubre tanto a e_1 como a e_2, y C expresa las condiciones particulares a cada objeto (en este caso, su propiedad de ser más pesado que el aire):

$\boldsymbol{P}_1$: $\forall X,Y.\ (o(X) \wedge o(Y)) \Rightarrow a(X,Y)$ (Existe una fuerza que atrae a los objetos (masivos) entre sí.)
$\boldsymbol{H}_1$: $\forall X.\ o(X) \Rightarrow a(tierra,X)$ (Los objetos masivos son atraídos hacia la tierra.)
$\boldsymbol{G}_1$: $(o(X) \wedge p(X)) \rhd a(tierra,X)$ (Los objetos más pesados que el aire tienden a caer hacia la tierra.)
$\boldsymbol{G}_2$: $(o(X) \wedge \neg p(X)) \rhd \neg a(tierra,X)$ (Los menos pesados que el aire tienden a no caer hacia la tierra.)
c_1: $p(piedra)$ (Esta piedra es más pesada que el aire.)
c_2: $\neg p(globo)$ (Este globo no es más pesado que el aire.)
e_1: $a(tierra,p)$ (Esta piedra cae hacia la tierra.)
e_2: $\neg a(tierra,g)$ (Este globo no cae hacia la tierra.)

En este caso, la teoría sistematiza mejor la evidencia, aunque aún está sujeta a excepciones (aviones en vuelo, por ejemplo). □

5 Comparación de teorías y la "lógica" del descubrimiento

En nuestra propuesta, el criterio de comparación de teorías, dentro del comportamiento de un programa, se basa en la relación de importancia epistémica de cada una. Es decir, las teorías a comparar deben tener una estructura epistémica común (comparten el mismo conocimiento), pero difieren en la importancia que le asignan a cada enunciado. Si las teorías a comparar no tienen la misma estructura, entonces es posible "igualarlas" agregando lo que le falta a cada una en el estrato de menor importancia epistémica. Un ejemplo que muestra cómo se comparan teorías es en el análisis de las distintas alternativas que surgen frente a una contrastación negativa.

Ejemplo 5.1: Sea la teoría $\boldsymbol{T} = \langle \{a, a \rhd b\}, \{\ \} \rangle$. Esta teoría predice b. Si b no es experimentalmente observada, o si hay evidencia de que $\neg b$ sucede en realidad, entonces podemos proponer por lo menos tres nuevas teorías:

1. En el primer caso proponemos $T_1=\langle\{a,\neg b,a \rhd b\},\{a \prec \neg b, a \prec (a \rhd b)\}\rangle$. De acuerdo con T_1, la predicción falló porque a no estaba adecuadamente justificada, pero $a \rhd b$ puede seguir siendo aceptada. Más aún, este estado de cosas puede incluso sugerir que $\neg a$ sea del caso, circunstancia que habrá que corroborar (ver más abajo).
2. Un segundo caso es proponer $T_2=\langle\{a,\neg b,a \rhd b\},\{a \prec b,(a \rhd b) \prec a\}\rangle$. En T_2, la causa de la refutación es la ley $a \rhd b$, la cual es negativamente contrastada por la evidencia de a y de $\neg b$, evidencia que puede seguirse manteniendo.
3. Otros casos más interesantes surgen de suponer una hipótesis particular c para proteger tanto a la evidencia como a la ley de la refutación. Un ejemplo semejante es una teoría $T_3=\langle\{a,\neg b,c,a \rhd b,(a \wedge c) \rhd \neg b\},\{\ \}\rangle$. Siguiendo a T_3, la ley $a \rhd b$ sistematiza parcialmente el dominio de la teoría, por lo que debe existir una ley más específica $(a \wedge c) \rhd \neg b$ que completa a la anterior en los casos en que se puede observar c. □

En este podemos observar situaciones similares a las vistas en la evolución de numerosos casos históricos (la radiación de fondo del universo, la deriva continental, la mecánica relativista y muchos más), donde se partió de una teoría T tradicionalmente aceptada, la cual fracasaba en algunos casos particulares. En esta situación, siempre se desea evitar llegar a la teoría T_2, dado que implica perder una ley científica frente a casos particulares. Históricamente la teoría, lejos de ser abandonada, fue protegida o bien rechazando los datos y buscando nuevos (caso T_1), o bien, cuando éstos datos se corroboraban (algo que llevaría al caso T_2 en forma inevitable), buscando condiciones particulares y nuevas leyes que la completaran (caso T_3). Es importante destacar que aquí existe un aspecto metodológico al que podemos denominar *metaestrategia*, dado que fuerza la elección de una determinada estrategia para defender al programa, generando nuevo conocimiento. Un grupo importante de casos de inferencia que llevan al descubrimiento de nuevo conocimiento se puede representar en nuestra formalización. En efecto, muchos ejemplos de razonamiento abductivo, razonamiento hipotético y cambio de teorías (Alchourón85) puede ser formalizado en nuestro sistema. Para ello se realiza el procedimiento de agregar el antecedente de la hipótesis a una estructura epistémica, asignándole la importancia del estrato de mayor importancia en la teoría. Si el consecuente del condicional forma parte del conjunto de conclusiones de la teoría ampliada, entonces el mismo queda justificado.

Ejemplo 5.2: Un esquema para representar una inferencia abductiva es el siguiente:

$$\frac{\begin{array}{l} b(t) \\ T \not\vdash b(t) \\ T \cup \{a(x)\} \vdash b(x) \end{array}}{a(t)}$$

es decir, *si b(t) es del caso, en T no es posible predecirlo, pero para todo X el agregado de a(X) permite predecir b(X) en T, entonces inferir a(t).* Una decisión importante para resolver aquí es qué importancia epistémica asignarle al enunciado inferido. Una estrategia posible es asignarle una importancia baja, e ir aumentándola en la medida en que el enunciado produzca contrastaciones exitosas. □

Ejemplo 5.3: Un esquema para representar una inferencia inductiva, dada una regularidad de casos, es:

$$\frac{a(t_1), a(t_2), \cdots, a(t_n)}{b(t_1), b(t_2), \cdots, b(t_n), \cdots, b(t_{n+m})}$$
$$a(x) \triangleright b(x)$$

es decir, *si cada vez que se encuentra a(t) se encuentra también b(t), entonces inferir una ley a(X) ▷ b(X).* □

Ejemplo 5.4: Un procedimiento para representar una inferencia hipotética condicional (problemática o contrafáctica) es el siguiente: Si en T un enunciado a se desconoce o es falso, y se plantea qué consecuencias tendría si a fuese verdadero (la *revisión* por a (Alchourón85)), entonces se agrega a a T con la máxima importancia epistémica, y se analizan las consecuencias. □

Si bien estos ejemplos son ingenuamente sencillos, muestran que es posible dar un aspecto formal a muchos de los patrones de inferencia que cubren los casos históricos más importantes. De todas maneras es importante destacar que el proceso de conjetura en el razonamiento científico es muy informado, y es rigurosamente analizado antes de ser sometido a consideración. Algunos de los aspectos indispensables a tener en cuenta para considerar adecuada una nueva ley son que la misma sea progresiva, es decir, que expanda el campo de aplicación que se tiene la teoría en forma consistente con el resto del conocimiento, que introduzca el menor número de cambios necesarios en el conocimiento anterior, y que no aumente el número de hipótesis *ad-hoc*. Esta consideración por parte de la comunidad constituye la realimentación crítica esencial en el proceso selectivo que referíamos más arriba.

6 Conclusiones

Se ha presentado una formalización de la presentación de Lakatos de los programas de investigación científica. Los mismos fueron caracterizados como una estructura epistémica compuesta por diversos tipos de conocimiento, una relación de preferencia epistémica que los estructura y un conjunto de patrones de inferencia. La misma utiliza como fundamento el desarrollo de los sistemas de razonamiento no monotónico de la inteligencia artificial. El sistema permite representar los distintos tipos de conocimiento existentes en una teoría científica, y permite implementar diversos aspectos del razonamiento como la inferencia inductiva, abducción, razonamiento hipotético, y demás. La relación de preferencia epistémica entre diversas piezas de conocimiento refleja la dimensión estratégica que tienen los programas de investigación. Un desarrollo futuro importante, entonces, es considerar a los distintos programas en competencia como un caso

de razonamiento multiagente. Otro tema a seguir desarrollando consiste en poder comparar teorías más allá de la relación de preferencia epistémica. Esto significa que deben existir elementos formales (sintácticos) que permitan dar mayor importancia a una teoría por su estructura que por su fundamento, como por ejemplo la especificidad, la presencia de subteorías preferidas, o el uso de mejor evidencia. Por último, queda también abierto el estudio de la *dinámica* de la preferencia epistémica, la cual refleja el cambio de estrategias en el comportamiento progresivo de los programas.

Agradecimiento: Algunas de las ideas aquí presentadas fueron discutidas con Fernando Tohmé, Guillermo Simari, Juan Manuel Torres y Jorge Roetti.

Referencias

C. Alchourón, P. Gardenfors, y D. Makinson. *On The Logic of Theory Change.* The Journal of Symbolic Logic, 50(2):510--530, 1985.

Roderick Chisholm. *Theory of Knowledge.* Pentice Hall, New Jersey, 1977.

Claudio Delrieux. *Incorporando Razonamiento Plausible en los Sistemas de Razonamiento Revisable.* Tesis de Magister en Ciencias de la Computación, Universidad Nacional del Sur, Departamento. de Ciencias de la Computación, 1995.

C. Delrieux y G. Simari. *Formalizing Plausible Reasoning.* XV International Conference of the Chilean Computer Society, págs 147-158, Arica, Chile, 1995.

Alicia Gianella. *Introducción a la Epistemología y a la Metodología de la Ciencia.* Universidad Nacional de La Plata, 1995.

Carl Hempel. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science.* The Free Press (NY), 1965.

C. Hempel y P. Oppenheim. *The Logic of Explanation.* Philosophy of Science, 15:135-175, 1948.

David Israel. *The Use of Logic in Knowledge Representation.* IEEE Computer, 16(10):37--42, 1983.

Gregorio Klimovsky. *Las Desventuras del Conocimiento Científico.* A-Z Editora, 1995.

Imre Lakatos. *Criticism and the Growth of Knowledge.* Cambridge University, 1970.

Imre Lakatos. *The Methodology of Scientific Research Programmes.* Cambridge University Press, 1978.

Ronald Loui. *Defeat Among Arguments: A System of Defeasible Inference.* Computational Intelligence, 3(3), 1987.

John McCarthy. *Epistemological Problems of Artificial Intelligence.* Proceedings of the Fifth International Joint Conference on Artificial Intelligence, págs 1038-1044, Morgan Kaufmann, Los Altos, CA, 1977.

John McCarthy. *Artificial Intelligence, Logic and Formalizing Common Sense.* Philosophical Logic and Artificial Intelligence (R. Thomason, editor), págs 161-190. Kluwer Academic Pub, 1989.

John McCarthy and Patrick Hayes. *Some Philosophical Problems from the Standpoint of Artificial Intelligence.* Machine Intelligence 4, págs 463-502. Edinburgh University Press, 1969.

David Poole. *On the Comparison of Theories: Preferring the Most Specific Explanation.* Proceedings of the Ninth International Joint Conference on Artificial Intelligence, págs 144-147, Morgan Kaufmann, Los Altos, CA, 1985.

Karl Popper. *The Logic of Scientific Discovery.* Hutchinson, London, 1959.

Raymond Reiter. *A Logic for Default Reasoning.* Art. Intelligence,13(1,2):81-132,1980.

Nicholas Rescher. *Hypotetical Reasoning.* North Holland, Amsterdam, 1974.

Nicholas Rescher. *Plausible Reasoning.* Van Gorcum, Dodrecht, 1976.

G. Simari y C. Delrieux. *Combinanado Plausibilidad y Razonamiento Revisable.* Jornadas Argentinas de Informática e Investigación Operativa, págs 99-110, 1994.