Una Interpretación Plausibilística de los Patrones Ampliativos de Inferencia

Claudio Delrieux

GIIA — Grupo de Investigación en Inteligencia Artificial
ICIC — Instituto de Ciencias e Ingeniería de Computación
Departamento de Ingeniería Eléctrica
Universidad Nacional del Sur
Alem 1253 — (8000) Bahía Blanca — ARGENTINA
e-mail:claudio@acm.org

Palabras Clave: Inteligencia Artificial, Razonamiento no Monotónico, Abducción, Inducción, Razonamiento Plausible

Resumen

El razonamiento no trivial a partir de premisas contradictorias es reconocido actualmente como una de las necesidades principales de los sistemas inteligentes, como por ejemplo en los sistemas expertos, planificadores y diagramadores autónomos de tareas y sistemas de diagnóstico. Esta necesidad surge de que la evidencia o conocimiento particular disponible en estos sistemas está normalmente expuesta a la presencia de información incompleta o potencialmente falible, como pueden ser generalizaciones accidentales, opiniones equivocadas o contrapuestas, información tendenciosa o deliberadamente falsa, errores operativos o de medición, etc.

El comportamiento esperado frente a una contradicción es aislarla del resto del conocimiento para no caer en la inconsistencia, o rechazar alguno de los términos contradictorios. En ambas situaciones se está estableciendo una inferencia más débil que la deducción, denominada *exducción*. En este trabajo buscamos aplicar la exducción no solamente para el razonamiento a partir de información contradictoria sino para el manejo de los patrones de inferencia ampliativos, como por ejemplo la abducción, la inducción, o el razonamiento no monotónico. La idea esencial es partir de una *estructura epistémica* que clasifica los diversos *tipos* de conocimiento por su justificación. Dentro de dicha estructura, una relación de *importancia epistémica* o plausibilidad permite asignar prioridades a cada uno de los elementos de la estructura epistémica. De esa manera, las conclusiones obtenidas por medio de inferencias ampliativas pueden representarse dentro de la estructura con la valoración o prioridad correspondiente.

1 Introducción

Uno de los problemas fundamentales de la Inteligencia Artificial, específicamente del área de representación de conocimiento y razonamiento (KR&R), es el manejo adecuado del conocimiento no universalmente válido, y los patrones de inferencia no deductivos normalmente asociados. En particular, el razonamiento no trivial a partir de reglas de inferencia con excepciones y de premisas contradictorias está siendo reconocido actualmente como una de las necesidades

principales en los sistemas inteligentes, como por ejemplo los sistemas expertos, planificadores y diagramadores de tareas y sistemas de diagnóstico. Esta necesidad surge de que tanto el conocimiento general como la evidencia o conocimiento particular disponible en estos sistemas está normalmente expuesta a información incompleta o potencialmente falible.

En estos casos se espera que los sistemas de razonamiento no se vuelvan trivialmente inconsistentes, dado que esto significaría que un solo dato erróneo arruina una base de conocimiento que puede ser valiosa. El comportamiento racional para un sistema inteligente en una situación de contradicción es aislarla del resto del conocimiento para no caer en la inconsistencia, o rechazar alguno de los términos contradictorios. Lo primero puede realizarse con una lógica paraconsistente, es decir, un sistema lógico más débil que la deducción [16, 22, 23]. Lo segundo se puede efectuar rechazando parte de la evidencia por medio de algún sistema extralógico [14, 33, 36, 41]. En ambas situaciones –el uso de una lógica paraconsistente o el uso de un sistema extralógico para aceptar la evidencia– se está estableciendo una inferencia más débil que la deducción. Es más, la clausura deductiva del conjunto de consecuencias es siempre un subconjunto de la clausura deductiva del antecedente. Por lo tanto, este tipo de inferencia no es ampliativa, utilizándose en la literatura filosófica el término *exducción* para denominarla.

En este trabajo buscamos aplicar la exducción no solamente para el razonamiento a partir de información contradictoria sino para el manejo de los patrones de inferencia ampliativos, como por ejemplo la abducción, la inducción, el razonamiento no monotónico o el razonamiento por analogía. La idea esencial es partir de una *estructura epistémica* que clasifica los diversos *tipos* de conocimiento por su justificación. Dentro de dicha estructura, una relación de *importancia epistémica* o plausibilidad permite asignar prioridades a cada uno de los elementos de la estructura epistémica. De esa manera, las conclusiones obtenidas por medio de inferencias ampliativas pueden representarse dentro de la estructura con la valoración o prioridad correspondiente. El conjunto de conclusiones que se justifica a partir de la estructura epistémica, entonces, puede encontrarse a partir de un sistema de razonamiento exductivo.

Dado que estas estructuras no son necesariamente consistentes, es posible encontrar subconjuntos consistentes máximamente plausibles. Cada subconjunto máximamente plausible de una estructura episteémica es una posible interpretación del conocimiento incompleto del dominio que se tiene en un contexto dado. Como normalmente pueden existir numerosas interpretaciones mútuamente inconsistentes, la semántica de nuestro sistema consiste en considerar como definitivo aquel conocimiento que que es común a todos los subconjuntos máximamente plausibles. Se presenta también un procedimiento de prueba para determinar cuáles sentencias están en estas condiciones, y se demuestra que dicho procedimiento es correcto y completo con respecto a la semántica.

La estructura de este trabajo es la siguiente. En la próxima sección introduciremos la caracterización formal de algunos de los patrones de inferencia ampliativa más usuales, como ser el razonamiento plausible [33, 34, 36], el razonamiento no monotónico [17, 19, 20, 31], el razonamiento inductivo [2, 35], y el razonamiento abductivo [3, 12, 24]. En la sección 3 se presentan las definiciones de la estructura epistémica y de un sistema de razonamiento exductivo basado en el sistema $\mathcal{P}$ de razonamiento plausible [6, 7] que se utiliza para justificar el conjunto de consecuencias de dicha estructura a partir de la importancia epistémica de las premisas. En la sección 4 estudiaremos la representación de los patrones ampliativos de inferencia mencionados en la sección 2 dentro del sistema $\mathcal{P}$, para utilizarlo, en la sección 5, en algunas aplicaciones y comparar los resultados obtenidos con otras propuestas. En la sección 6, por último, se discutirán las conclusiones y los comentarios finales.

2 Inferencia ampliativa

La inferencia puede entenderse como una relación entre un conjunto de sentencias (denominado *antecedente*, del cual cada uno de sus miembros son *premisas*) y una sentencia (denominada *consecuente*), de modo tal que la valoración epistémica o soporte asertivo del consecuente queda determinado por la valoración epistémica o soporte asertivo del antecedente [29]. De esa forma, la inferencia se interpreta como una relación de soporte asertivo. Una inferencia se denomina *ampliativa* si el contenido informacional del consecuente excede al del antecedente, y se denomina *monotónica* o *aditiva* si pueden agregarse premisas arbitrariamente, sin que se modifique el soporte asertivo del consecuente.

Por *deducción* entendemos una inferencia que procede en forma demostrativa, es decir, preservando la verdad. La inferencia deductiva es *analítica*, es decir, las conclusiones que se siguen de un conjunto de premisas están ya contenidas en dichas premisas. Por ejemplo, si conocemos que *opus* es un pingüino, y que los pingüinos son aves, entonces está implícito el hecho de que *opus* es un ave. Por lo tanto, la inferencia deductiva es *correcta*, en el sentido de que no puede generar una conclusión incorrecta. Ésto significa que la deducción debe ser monotónica, dado que el agredado de premisas debe seguir infiriendo una conclusión correcta. Sin embargo, en cierto sentido la inferencia deductiva no es informativa. Por mucho que se aplique, no generará conocimiento nuevo. Por dicha razón la deducción es no ampliativa.

Mucho se ha ejemplificado cómo en la vida real un agente necesita inferir cosas que aún con una inferencia deductiva ideal (omnisciente) no pueden inferirse. Por ejemplo, inferir que normalmente un ave es voladora. Este tipo de razonamiento es *sintético*, es decir, va "más allá" de sus premisas. Por dicha razón este tipo de inferencia es *ampliativa*. La inferencia ampliativa, sin embargo, no puede ser correcta, dado que pueden ocurrir casos en que las premisas son verdaderas pero la conclusión no, como en el caso de inferir que *opus* es volador. La inferencia ampliativa se da siempre dentro de un *contexto* normalmente cerrado, dentro del cual, pese a no ser correcta, es indispensable. La modificación del contexto puede, entonces, llevar a abandonar conclusiones que antes eran aceptadas. Por dicha razón, la inferencia ampliativa es no monotónica. Es importante notar que la conversa no vale necesariamente, es decir, puede existir inferencia no monotónica que sea no ampliativa. La exducción es precisamente un ejemplo.

En esta sección presentaremos una introducción a los patrones de inferencia ampliativa más usuales, y de mayor utilidad en la inteligencia artificial, como ser el razonamiento no monotónico, la inducción, la abducción, y el razonamiento plausible. Cada uno de estos patrones cumple con objetivos diferentes y en cierto punto complementarios. El razonamiento no monotónico busca extender el conjunto de conclusiones de una teoría por medio de un conjunto de reglas tentativas conocidas y aceptadas *a priori*. El razonamiento inductivo, en cambio, genera este tipo de reglas a partir de la observación de un conjunto de casos particulares en los que se observa una regularidad. El razonamiento plausible busca extender el dominio de una teoría por medio de la aceptación de información u observaciones que provienen de diversas fuentes, incluída la formación de hipótesis y la conjetura. La abducción busca explicar un resultado no predicho por una teoría en función de alguna observación que la justifique y que ha sido pasada por alto.

2.1 Razonamiento no monotónico

El razonamiento no monotónico, como vimos, surge a partir de procedimientos de inferencia no demostrativos. En las presentaciones usuales, la inferencia no monotónica se produce a partir del uso de premisas útiles pero no siempre válidas en un esquema de inferencia que, por todo lo demás, es idéntico a una inferencia deductiva, particularmente a *modus ponens*. Estas premisas tentativas intentan representar una forma contextual de razonamiento que es sin duda útil e indispensable en tanto que dicho contexto se mantenga fijo (condición de *ceteris paribus*). En un sentido amplio, la premisa *Normalmente, cuando a sucede, entonces b también sucede* intenta representar la predisposición a realizar la inferencia no monotónica de b cuando en el contexto de la teoría es posible demostrar a pero no es posible demostrar $\neg b$. Dicha inferencia será no monotónica porque, en cuanto b pueda genuinamente demostrarse en el contexto, entonces la conclusión de la inferencia queda *bloqueada*.

El primer objetivo de estos sistemas fue establecer la posiblidad de representar patrones de inferencia que nos permitan cerrar el dominio de una teoría incompleta y poder extraer información negativa de la misma. Por ejemplo, una teoría extensional (o *base de datos*) de una línea aérea es una colección de aserciones, cada una de las cuales representa las características de cada vuelo de la compañía. En estas condiciones es posible responder afirmativamente cuando se consulta acerca de un determinado vuelo y el mismo pertenece a la base de datos. Sin embargo *no es posible* responder negativamente cuando la consulta no pertenece, dado que la teoría es incompleta. Para poder completar la teoría es necesario agregar la negación de las aserciones de todos los vuelos de que la compañía no dispone (lo cual es absurdo), o bien cerrar el dominio con un esquema que represente el hecho de que *si una aserción no pertenece a la base de datos, entonces se puede inferir su negación*. Esto es lo que se conoce como *suposición de mundo cerrado* [30].

Otra característica que se busca representar en estos sistemas es la ocurrencia de propiedades *prima facie* (o prototipos) en determinados individuos de la teoría. Por ejemplo, una teoría acerca de características típicas de las aves puede incluír la regla tentativa de que *normalmente las aves vuelan*. Esto se puede representar con un "condicional derrotable" según el cual, cada vez que existe un individuo X de quien se conoce la propiedad de ser ave, y no hay razones para abstenerse de inferir que X posee la capacidad de volar, entonces se infiere que el individuo X posee la capacidad de volar. Una razón adicional proviene de la semidecidibilidad de las teorías. La búsqueda de demostraciones para sentencias que no son teoremas puede llevar a procesos computacionalmente infinitos. Por lo tanto, un esquema no monotónico para rodear el obstáculo de la semidecidibilidad es utilizar la *negación por falla* [5, 37], según la cual, si una sentencia s no puede demostrarse, entonces se infiere $\neg s$.

Los dispositivos formales utilizados para representar este tipo de premisas, asociados a la inferencia no monotónica, utilizan implicaciones materiales en un lenguaje modal [19, 20], o bien esquemas de regla en el metalenguaje [31], por medio de operadores condicionales (metalingüísticos) [8, 21], por medio de relaciones de "anormalidad" debilitando implicaciones materiales en teorías incompletas [17] o bien como relaciones metalingüísticas entre conjuntos de literales [15, 25]. En este trabajo adoptamos este último esquema, es decir, la expresión $\alpha \succ\!\!\!- \beta$ es un *condicional derrotable* o implicación *prima facie* para representar el conocimiento de que las razones para aceptar α proveen razones para aceptar tentativamente a β, y tanto α como β son conjuntos de literales. Puede establecerse una relación de *consecuencia tentativa* $\mid\!\sim$ extendiendo la relación de consecuencia clásica $\vdash$ de modo tal que se utilicen estos condicionales derrotables como implicaciones materiales en una deducción clásica. De acuerdo a esta repre-

sentación, el esquema de inferencia no monotónica es:

$$\frac{\begin{array}{l} a(t) \\ a(X) \succ\!\!- b(X) \end{array}}{b(t),} \tag{1}$$

es decir, *si es posible observar que $a(t)$, y entre nuestros condicionales derrotables se sostiene que $a(X) \succ\!\!- b(X)$, entonces inferir $b(t)$.*

2.2 Inducción

El objetivo de la inferencia inductiva o inducción es encontrar una regla general a partir de un conjunto de casos particulares, instancias o ejemplos [2]. La misma es de importancia central en temas de la inteligencia artificial, como en aprendizaje, reconocimiento de patrones o búsqueda en juegos, y también en la filosofía y la teoría de la ciencia. Una manera de expresar formalmente el objetivo de la inducción es decir que busca una *abstracción* de la cual los casos particulares sean instancias. Esta abstracción constituye una clase de equivalencia (probablemente parcial) en el conjunto de casos particulares.

Un método fundamental de inferencia inductiva es la búsqueda sistemática, de modo de encontrar lo antes posible una regla que sea consistente con el conjunto de ejemplos. Para que este método sea computable, debe existir un algoritmo que genere las reglas de manera sistemática, y que chequee su consistencia con los ejemplos. Reynolds [35] investigó la estructura algebraica del conjunto de fórmulas de un lenguaje de primer orden bajo la relación de sustitución, estableciendo que dicha estructura conforma un reticulado, y que existe una forma de computación simple para encontrar la menor instancia más general. Un esquema adecuado para representar una inferencia inductiva, dada una regularidad de casos particulares, es

$$\frac{\begin{array}{l} a(t_1), a(t_2), \cdots, a(t_n) \\ b(t_1), b(t_2), \cdots, b(t_n), \cdots, b(t_{n+m}) \end{array}}{a(X) \succ\!\!- b(X),} \tag{2}$$

es decir, *si en nuestra base de conocimientos cada vez que se encuentra $a(t)$ se encuentra también $b(t)$, entonces inferir $a(X) \succ\!\!- b(X)$*, donde la relación $a(X) \succ\!\!- b(X)$, como vimos, establece el condicional derrotable o implicación *prima facie* de que las razones para aceptar $a(X)$ proveen razones para aceptar $b(X)$.

2.3 Abducción

En términos generales debidos a Peirce y otros filósofos, la abducción es el proceso de inferencia que va de las observaciones a las explicaciones dentro de un contexto o teoría general. Es decir, la inferencia abductiva busca sentencias (denominadas *explicaciones*) que, agregadas a la teoría, implican deductivamente a las observaciones. Muchas veces, sin embargo, existen varias explicaciones posibles para una observación, por lo que el "arte" de la inferencia abductiva es encontrar la explicación que sea "mejor" en algún sentido. La abducción tiene una importancia central en muchos sistemas de inteligencia artificial, como por ejemplo en los sistemas expertos de diagnóstico, la detección de fallas y el razonamiento causal [12, 24, 32].

Debe entenderse, dado que es un error muy común, que explicar no es lo mismo que implicar. Tomar mate con un engripado *explica* un contagio, pero no lo *implica*. Este punto sin duda

no ha recibido toda la atención debida, dado que el estudio de la abducción se realizó en general en el contexto de las teorías monotónicas, por ejemplo, dentro de la teoría de la ciencia, en el paradigma hipotético-deductivo [11]. Cuando existen reglas no monotónicas, pueden ocurrir ejemplos como el mencionado, en el que una buena explicación normalmente es una mala predicción.

Un esquema para representar una inferencia abductiva es el siguiente:

$$\frac{\begin{array}{l} b(t) \\ \mathcal{T} \not\mid\sim b(t) \\ \mathcal{T} \cup a(X) \mid\sim b(X) \end{array}}{a(t),} \tag{3}$$

es decir, *si se produce la observación de un hecho $b(t)$ el cual no es predicho en nuestra teoría $\mathcal{T}$, pero encontramos que a(X), agregado a la teoría, es una explicación de b(X), entonces inferimos a(t).*

2.4 Razonamiento plausible

El razonamiento plausible, como mecanismo agregado a un sistema lógico, tiene como tarea principal extender el dominio de una teoría incorporando información o conocimiento que provenga de fuentes de información potencialmente falible. En este trabajo consideraremos que la información plausible que se puede aceptar es únicamente aquello que se puede observar, es decir, literales de base. La justificación filosófica de ésto es un tanto extensa y puede consultarse en [6], pero intuitivamente se puede comprender que el proceso de aceptar información externa consiste en incorporar observaciones, mediciones, y eventualmente opiniones que uno mismo formula o acepta de otros agentes de confianza. Es importante considerar que estos criterios de información pueden provenir también de conjeturas o de procesos de formación de hipótesis (por ejemplo por abducción), pero siempre restringidos a considerar literales, dado que otros casos, por ejemplo conjeturar reglas, quedan dentro del ámbito del razonamiento inductivo.

Uno de los problemas que puede ocurrir al incorporar conocimiento plausible, es que éste puede llevar a conclusiones contradictorias. Una estrategia para restaurar la consistencia consiste en abandonar la mínima cantidad conocimiento plausible, y en particular aquel que por alguna razón signifique la menor pérdida de conocimiento útil. Cada premisa plausible, entonces, estará acompañada de un índice que determine su importancia desde un punto de vista epistemológico. Por ejemplo, si el agente (o criterio de información) i nos informa que el hecho (literal de base) $a(X)$ ocurre, entonces se representa como $< a(X) >_i$. Entonces, la plausibilidad de una premisa es una medida de su grado de soporte asertivo, es decir, de su valoración epistemológica [4].

Este es uno de los sentidos en los que el razonamiento plausible se diferencia del razonamiento probabilístico: la razón de aceptar una opinión no proviene de la opinión en sí misma sino de quién la provee. Es decir, la plausibilidad como valoración epistemológica de una sentencia es *externa* al contenido intensional de la misma. Por ejemplo, una amiga arroja un dado y nos dice que salió el 4. La probabilidad de dicho evento es $\frac{1}{6}$. Pero nosotros no tenemos razones para desconfiar de nuestra amiga, por lo que la plausibilidad que para nosotros tiene el evento es mucho mayor que $\frac{1}{6}$.

Restaurar la consistencia consistirá, entonces, en abandonar la menor cantidad de premisas plausibles necesarias, y de entre ellas las que menor plausibilidad tengan. Este tipo de criterios

son comunmente utilizados en epistemología y en teoría de la ciencia [10, 13, 27]. Por ejemplo, cuando un experimento contradice una teoría firmemente establecida, se descarta el resultado del experimento, pues su valor epistemológico es menor que el valor epistemológico de la teoría. Algo similar ocurre con el razonamiento con sentido común. Cuando recibimos opiniones contradictorias, nuestra tendencia es la de rechazar implícitamente la opinión vertida por aquella persona en la que tenemos menor confianza. Es decir, las opiniones se consideran no por su contenido sino por su fuente de información.

3 Una arquitectura epistémica basada en plausibilidad

En esta sección presentaremos una "arquitectura epistémica", es decir, una estructura de representación de conocimiento establecida en función de un modelo específico de razonamiento. La misma se basa en la relación de preferencia epistémica para realizar exducción, es decir, para descartar de un antecedente inconsistente las premisas de menor importancia. Muchos aspectos de este sistema están basados en el sistema $\mathcal{P}$ de razonamiento plausible [6, 7], aunque con una motivación diferente.

El conocimiento del sistema se representa en un un lenguaje de primer orden deductivamente cerrado. Utilizaremos la convención de nombrar con letras minúsculas italizadas (por ejemplo $a(X)$) para referirnos a sentencias arbitrarias en dicho lenguaje, mayúsculas caligráficas (por ejemplo $\mathcal{K}$) para referirnos a conjuntos de sentencias generales (nonground), y con mayúsculas italizadas (por ejemplo C) para referirnos a conjuntos de sentencias particulares (ground). Como vimos, 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 derrotables. Por ejemplo, la expresión $a(X) \succ\!\!\!- b(X)$ expresa que *"La disposición de aceptar $a(X)$ es una razón para aceptar tentativamente $b(X)$"*. Las sentencias que representan conocimiento tentativo particular, que asume la forma de evidencia plausible, 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, inferencia abductiva, inferencia no monotónica, etc.).

En nuestra definición, una *arquitectura epistémica* $\mathcal{T}$ está constituída por la unión de enunciados pertenecientes a los siguientes conjuntos de conocimiento:

- $\mathcal{K}$, *conocimiento lógico-matemático* y las *definiciones*, que son, obviamente, deductivamente válidas;
- $\mathcal{G}$, *conocimiento tentativo* representado como condicionales derrotables o implicaciones *prima facie* que surgen como abstracción de un conjunto razonablemente grande de casos particulares;
- E, la *evidencia* es el conjunto de conocimiento particular disponible, el cual también es deductivamente válido;
- P, la *información plausible* representada como conocimiento particular *prima facie*, que proviene de alguna fuente de información (medición, opinión), o de un proceso de formación de hipótesis (como consecuencia de una inferencia abductiva, inferencia no monotónica, etc.).

DEFINICIÓN **3.1** *Dado un contexto $< \mathcal{K}, E >$, (el conocimiento lógico-matemático, las definiciones y la evidencia), una* **Estructura Epistémica** $\mathcal{E}_{\mathcal{K},E}$ *es una estructura de conocimiento*

$\mathcal{E}_{\mathcal{K},E} =< \mathcal{G}, P >$, *donde* $\mathcal{G}$ *es un conjunto finito de condicionales de la forma* $\alpha \succ\!\!\!- \beta$, *y* C *es un conjunto de información plausible representado como literales de base de la forma* $< l >_i$. *Cuando el contexto quede claramente definido, nos referiremos a una estructura epistémica simplemente como* $\mathcal{E}$.

Uno de los elementos esenciales en nuestra formalización consiste en representar una relación de preferencia epistémica dentro de los elementos de conocimiento en una estructura. Los únicos conjuntos de conocimiento firme, es decir, aquellas piezas de conocimiento que no pueden ser en principio cuestionadas, son las pertenecientes al contexto $\mathcal{K}$ y E.

DEFINICIÓN **3.2** *Dada una estructura epistémica* $\mathcal{E}_{\mathcal{K},E}$ *en un contexto* $< \mathcal{K}, E >$, *una* **Teoría Plausible** $\mathcal{T}$ *es un par* $\mathcal{T} =< \mathcal{E}_{\mathcal{K},E}, \prec >$, *donde* $\prec$ *es un orden parcial sobre los enunciados de* $\mathcal{E}$, *llamado relación de* **Preferencia Epistémica** *o plausibilidad. Podemos considerar que el conocimiento deductivamente válido del contexto conforma un elemento* $\mathcal{E}_\top$ *tal que* $\forall \alpha \in \mathcal{E}.\alpha \prec \mathcal{E}_\top$, *y que todo conocimiento que no tenga una importancia epistémica destacable conforma otro elemente* $\mathcal{E}_\bot$ *tal que* $\forall \beta \in \mathcal{E}.\mathcal{E}_\bot \prec \beta$. *De esa manera, en una teoria plausible* $\mathcal{T}$, *todas las piezas de conocimiento de su estructura epistémica* $\mathcal{E}$ *quedan reticulados bajo la relación* $\prec$.

Dado que las teorías plausibles no son necesariamente consistentes, la idea esencial es que las conclusiones plausibles de una teoría son consecuencia de sus subteorías consistentes (con respecto al contexto) máximamente plausibles. Por ello es necesario poder comparar la plausibilidad de dichas subteorías.

DEFINICIÓN **3.3** *Dada una teoría* $\mathcal{T}$ *y una* subteoría $T \subseteq \mathcal{T}$, *la* **Importancia Epistémica** *de* T *dada* $\mathcal{T}$, *denotada por* $T_\prec$ *se define como el conjunto de cotas inferiores de* T *bajo la relación* $\prec$ *de preferencia epistémica:* $T_\prec = \{\alpha \in T |\ \nexists \beta \in T.\beta \prec \alpha\}$. *Dadas dos subteorías* T_1 *y* T_2, *diremos que* T_1 *es epistémicamente más importante que* T_2 *(denotado como* $T_2 \prec T_1$*) si y sólo si cada enunciado en* T_1 *es al menos tan importante en* $\mathcal{T}$ *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 $\mathcal{T}$, ¿cuál es el subconjunto consistente de enunciados de $\mathcal{T}$ de mayor importancia epistémica? La semántica aquí propuesta consiste en considerar que conforman la intersección de todos los conjuntos consistentes (con respecto al contexto) de mayor plausibilidad, generados bajo distintas extensiones lineales de $\prec$. Esta manera de razonar constituye un acercamiento escéptico, es decir, ante la imposibilidad de establecer una preferencia entre dos o más piezas de conocimiento mútuamente inconsistentes (con respecto al contexto), entonces se abstiene de aceptarlas a todas.

DEFINICIÓN **3.4** *Dada una teoría* $\mathcal{T} =< \mathcal{E}, \prec >$, *una* **Extensión Lineal** l *de* $\prec$ *es una relación que contiene a* $\prec$ *y que induce un orden lineal en* $\mathcal{E}$.

EJEMPLO **3.1** *Supongamos que tenemos los enunciados* $\mathcal{E} = \{a, b, c\}$ *y que la relación de preferencia en* $\mathcal{E}$ *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 **3.5** *Sea el operador de consecuencia* $\mid\sim$ *que puede incorporar subconjuntos de una estructura epistémica como parte de su antecedente para expresar relaciones de consecuencia que resultarían si los enuciados de dichos subconjuntos fueran utilizados por las reglas de inferencia del operador de consecuencia deductiva clásico, particularmente los miembros de* $\mathcal{G}$ *como implicaciones materiales y los miembros de* C *como literales. En dicho caso, sea una teoría* $\mathcal{T} =< \mathcal{E}, \prec >$ *en un contexto* $< \mathcal{K}, E >$*, y una extensión lineal* l *de* $\prec$*, una* **Subteoría Máximamente Consistente** *(SMC) de* $\mathcal{T}$ *(con respecto al contexto* $< \mathcal{K}, E >$*) es un subconjunto* $\mathcal{E}^l$ *de la estructura epistémica* $\mathcal{E}$ *que satisface:*

1. $\mathcal{E}^l \subseteq \mathcal{E}$,
2. $(\mathcal{E}^l \cup \mathcal{K} \cup \mathcal{P}) \not\mid\sim \bot$,
3. $\forall \alpha \in \mathcal{E}^l . \forall \beta \in (\mathcal{E}/\mathcal{E}^l) . \beta \prec \alpha$,
4. $\nexists \mathcal{E}' . \mathcal{E}^l \subset \mathcal{E}' \subseteq \mathcal{E}, (\mathcal{E}' \cup \mathcal{K}) \not\mid\sim \bot$,

es decir, una SMC $\mathcal{E}^l$ *es un subconjunto consistente de la estructura epistémica* $\mathcal{E}$ *(condiciones 1 y 2), tal que no existe ninguna pieza de conocimiento en* $\mathcal{E}$ *no incluída en* $\mathcal{E}^l$ *que sea estrictamente más importante epistémicamente que todas las piezas de conocimiento en* $\mathcal{E}^l$ *(condición 3), y no existe ninguna subteoría consistente de* $\mathcal{T}$ *que incluya propiamente a* $\mathcal{E}^l$ *(condición 4). La intersección de todos los SMC de* $\mathcal{T}$ *es un subconjunto de* $\mathcal{E}$*. Si consideramos a los condicionales en dicha intersección como implicaciones materiales, y a la información plausible como literales, obtenemos la* **Subteoría Escéptica** $\mathcal{T}_{\not\sim}$ *de* $\mathcal{T}$ *(con respecto al contexto* $< \mathcal{K}, E >$ *y a la relación de importancia epistémica* $\prec$*). Dicha subteoría está dentro del lenguaje de la lógica clásica. El conjunto* C *de* **Conclusiones** *(predicciones o explicaciones) de una teoría plausible* $\mathcal{T}$*, entonces, es la clausura deductiva de su subteoría escéptica junto con el contexto, es decir,* $C = Th(\{\mathcal{K} \cup E \cup \mathcal{T}_{\not\sim}\})$.

Cada subteoría máximamente consistente (SMC) de una estructura epistémica es una posible interpretación del conocimiento incompleto del dominio que se tiene en la teoría plausible en un contexto dado. Como normalmente pueden existir numerosas interpretaciones mútuamente inconsistentes, la subteoría escéptica es aquella parte que es común a todas las subteorías consistentes. Un procedimiento de prueba correcto y completo para determinar si una sentencia dada está en el conjunto de conclusiones de una teoría es el siguiente:

DEFINICIÓN **3.6** *Dada una teoría* $\mathcal{T} =< \mathcal{E}_{\mathcal{K},E}, \prec >$ *y una consulta* q *tal que ni* $\mathcal{K} \cup E \vdash q$ *ni* $\mathcal{K} \cup E \vdash \neg q$*. Entonces definimos:*

(Fundamento) *q tiene fundamento si existe un conjunto de fundamento $\mathcal{E}_f \subseteq \mathcal{E}$, tal que $\mathcal{E}_f \cup \mathcal{K} \cup E \mid\sim q$.*

(Duda) *q está en duda si existe un conjunto de duda $\mathcal{E}_d \subseteq \mathcal{E}$, tal que $\mathcal{E}_d \cup \mathcal{K} \cup E \mid\sim \neg q$.*

(Aceptación) *q es aceptado si tiene fundamento y no está en duda, o bien si existe un conjunto de fundamento $\mathcal{E}_f$ tal que para cualquier conjunto de duda $\mathcal{E}_d$ se cumpla que $\mathcal{E}_d \prec \mathcal{E}_f$, es decir, la importancia epistémica del conjunto de fundamento es mayor que la del conjunto de duda.*

El siguiente teorema muestra que el procedimiento de aceptación descripto en la definición 3.6 es correcto y completo con respecto a la definición 3.5 del conjunto de conclusiones.

TEOREMA 1 *Dada una teoría* $\mathcal{T} =< \mathcal{E}_{\mathcal{K},E}, \prec >$, *entonces* q *es aceptada con fundamento* $\mathcal{E}_f$ *tal que* $\emptyset \subset \mathcal{E}_f \subseteq \mathcal{E}$, *si y solo si* q *pertenece al conjunto* C *de conclusiones.*

DEMOSTRACIÓN
$\Leftarrow$
Si q *pertenece al conjunto* C *de conclusiones, entonces* $\mathcal{K} \cup E \cup \mathcal{T}_{\not\perp} \models q$, *y por lo tanto,* $\mathcal{K} \cup E \cup \mathcal{E}^l \models q$, *para toda SMC* $\mathcal{E}^l$ *que se obtienen en toda extensión lineal* l *de la relación* $\prec$. *Luego, en toda MCS* $\mathcal{E}^l$ *de* $\mathcal{T}$ *(con respecto al contexto* $< \mathcal{K}, E >$*) existe algún conjunto de fundamento* $\mathcal{E}^l_s$ *tal que* $\mathcal{K} \cup E \cup \mathcal{E}^l_s \models q$. *En dicho caso, supongamos que existe un conjunto de duda* $\emptyset \subset \mathcal{E}_d \subseteq \mathcal{E}$ *tal que, en conjunción con el contexto* $< \mathcal{K}, E >$ *impliquen lógicamente a* $\neg q$. *Si no existese un conjunto de duda* $\mathcal{E}_d$ *de tales características, entonces* q *estaría fundamentada y no estaría en duda, por lo que* q *sería aceptada (Q.E.D.). Si existe un conjunto de duda* $\mathcal{E}_d$, *entonces, dado que* $\mathcal{T}_{\not\perp}$ *es consistente por hipótesis, entonces debe ser que* $\mathcal{E}^l_d \prec \mathcal{E}^l_s$ *en toda extensión lineal* l *de* $\prec$, *y en concecuencia, en la relación* $\prec$ *misma se debe cumplir* $\mathcal{E}_d \prec \mathcal{E}_s$. *En dicho caso* q *es aceptada, dado que su conjunto de fundamento es de mayor importancia epistémica que su conjunto de duda, en toda extensión lineal de la relación* $\prec$. *(Q.E.D.).*
Esto completa la primera mitad de la demostración.

$\Rightarrow$
Supongamos que q *es aceptada con un conjunto de fundamento* $\mathcal{E}_s$ *tal que* $\emptyset \subset \mathcal{E}_s \subseteq \mathcal{T}_{\not\perp}$. *En dicho caso, supongamos que existe un conjunto de duda* $\mathcal{E}_d$ *tal que* $\emptyset \subset \mathcal{E}_d \subseteq \mathcal{E}$ *y que junto con* $\mathcal{K}$ *y* E *impliquen lógicamente a* $\neg q$. *Nuevamente, si no existiera un conjunto de duda* $\mathcal{E}_d$, *entonces* $\mathcal{E}_s$ *sería consistente con el contexto* $< \mathcal{K}, E >$, *y, por lo tanto,* $(\mathcal{K} \cup E \cup \mathcal{T}_{\not\perp})$ *implicaría lógicamente a* q *sin contradicción (Q.E.D.). Si existe un conjunto de duda* $\mathcal{E}_d$, *entonces, dado que por hipótesis* q *es aceptada, entonces se debe cumplir que* $\mathcal{E}_d \prec \mathcal{E}_s$. *En dicho caso, de acuerdo a la definición 3.3, en toda extensión lineal* l *se debe cumplir la condición* $\mathcal{E}^l_d \prec \mathcal{E}^l_s$. *Entonces se sigue que* $\mathcal{E}_s$ *debe pertenecer a todas las MCS de* $\mathcal{E}$ *(con respecto al contexto* $< \mathcal{K}, E >$*), y consecuentemente a la subteoría escéptica* $\mathcal{T}_{\not\perp}$ *que es la intersección de todas dichas SMC. Por fin, la subteoría escéptica de* $\mathcal{T}$, *junto con el contexto, deben implicar lógicamente a* q *sin contradicción (Q.E.D.).*

El procedimiento descripto en la definición 3.6, inspirado en el razonamiento escéptico de Wagner [41], es computacionalmente tratable, dado que está basado en demostrabilidad por *backward chaining* aplicada recursivamente. En consecuencia, una implementación de dicho procedimiento en PROLOG es directa.

4 Inferencia ampliativa en la arquitectura epistémica

En esta sección presentaremos algunos ejemplos de los esquemas de inferencia ampliativa tal como son representados en la estructura epistémica presentada en la sección anterior. Para simplificar la notación en dichos ejemplos, utilizaremos la notación de Geffner para los condicionales [9]. Cada condicional en $\mathcal{G}$, entonces, es indexado con un número que lo representa, de modo que el condicional $\alpha \succ\!\!\!\!\!\!\sim_{\,i}\ \beta$ es representado como δ_i.

4.1 Razonamiento no monotónico

Como es posible observar, en la definición 3.5 de consecuencia de una teoría plausible y en su contraparte operacional en la definición 3.6 está implícito un esquema de inferencia no monotónica similar al esquema 1, al considerar los condicionales derrotables como implicaciones materiales en la regla *modus ponens*. Por lo tanto, la arquitectura epistémica tal como fue presentada en la sección anterior efectúa una forma de razonamiento no monotónico basada en la preferencia epistémica entre condicionales derrotables.

Se puede demostrar que, descartando o desconociendo la relación de preferencia entre condicionales, el conjunto de conclusiones obtenido a partir subteoría escéptica coincide con las conclusiones de una teoría no monotónica de McDermott y Doyle [19] o una teoría autoepistémica [20]. Por otro lado, utilizar cada una de las subteorías consistentes máximamente plausibles obtiene conjuntos de conclusiones mutuamente incompatibles idénticos a las extensiones múltiples producidas por el razonamiento *Default* de Reiter [31]. Esta manera constituye una forma "crédula" de razonar, y cada uno de estos conjuntos de conclusiones es una posible visión del mundo compatible con la teoría plausible. Sin embargo, el uso de la relación de preferencia permite romper con la dualidad del razonamiento escéptico *versus* el crédulo.

EJEMPLO **4.1** *Supongamos que en nuestra teoría tenemos*

$E = \{a, \neg(c \wedge d)\}$,
$\mathcal{G} = \{a \succ_1 b, b \succ_2 c, a \succ_3 d\}$ *y*
$\{\delta_1 \prec \delta_3, \delta_2 \prec \delta_3\}$.

En este caso es fácil ver a partir de las definiciones dadas en la sección anterior que d *es conclusión de nuestra teoría, dado que si utilizamos a* $\{\delta_3\}$ *como fundamento, es posible concluir* d *en forma consistente a partir del contexto, y dicho fundamento es preferible al conjunto de duda* $\{\delta_1, \delta_2\}$. *Es importante observar que* b *no es conclusión de la teoría, dado que su fundamento* δ_1 *no es comparable con su conjunto de duda* $\{\delta_2, \delta_3\}$. *Si en* $\prec$ *se agregara* $\delta_2 \prec \delta_1$, *entonces* b *sería también conclusión de la teoría, y si en cambio en* $\prec$ *se agregara* $\delta_1 \prec \delta_2$, *entonces* $\neg b$ *sería la nueva conclusión (al utilizarse* δ_2 *en forma contrapositiva).*

Estos resultados son similares a los producidos con sistemas de razonamiento no monotónico con preferencias, como por ejemplo circunscripción priorizada [18, 28], pero con un mejor comportamiento desde el punto de vista computacional. Al mismo tiempo, es fácil ver que este modelo de razonamiento no monotónico no cae en las "trampas" de la ambigüedad, como algunas formas del razonamiento basado en redes de herencia [39].

EJEMPLO **4.2** [**Ambigüedades en Cascada** *(Horty et. al., 1987)*]
El conocimiento general que el sistema tiene acerca de algunas actitudes políticas es el siguiente:

Los republicanos no son pacifistas	$r \succ_1 \neg p$
Los cuáqueros son pacifistas	$c \succ_2 p$
Los republicanos son fanáticos del futból	$r \succ_3 ff$
Los fanáticos del futból son belicistas	$ff \succ_4 b$
Los pacifistas no son belicistas	$p \succ_5 \neg b$
Nixon es republicano	r
Nixon es cuáquero	q

¿Qué puede concluír el sistema con respecto al belicismo de Nixon? En particular tenemos las siguientes líneas de razonamiento:

$\{\delta_2, \delta_5\} \cup \mathcal{K} \cup E \vdash \neg b$,
$\{\delta_1, \delta_5\} \cup \mathcal{K} \cup E \vdash b$, *y*
$\{\delta_3, \delta_4\} \cup \mathcal{K} \cup E \vdash b$.

La conclusión aceptada dependerá, entonces, de la preferencia en la relación de preferencia de alguno de los tres conjuntos $\{\{\delta_2, \delta_5\}, \{\delta_1, \delta_5\}, \{\delta_3, \delta_4\}\}$ *con respecto a los otros dos.*

4.2 Inducción

A diferencia del razonamiento no monotónico, en el cual se encuentra una conclusión plausible, en el razonamiento inductivo no se encuentra una conclusión, sino un condicional derrotable, el cual, a su vez, podrá ser utilizado por otro esquema de inferencia para producir conclusiones. De esa manera, al realizar un razonamiento inductivo dentro de nuestra arquitectura epistémica, nos enfrentamos a un nuevo problema, consistente en asignar un índice de importancia epistémica a dicho condicional. Esta dificultad forma parte de una problemática más general, estudiada en la teoría de la ciencia como "el problema de la inducción", especialmente por Carnap y Popper [26, 27], dentro de la búsqueda de un criterio estadístico de confirmación de teorías científicas.

Si bien para muchos se llegó a la conclusión de que la vía estadística es inadecuada para resolver el problema de la inducción y la confirmación [13], algunos de los criterios propuestos parecen tener sentido en el caso que nos ocupa. Por ejemplo, en el esquema de inferencia inductiva mostrado más arriba (ver 2), la cantidad n de casos coincidentes queda indeterminada, pudiendo ser 1. Ésto llevaría a una especie de "explosión inductiva" (como la que ocurre en el sistema PI de Thagard [38]) en la cual cada vez que dos predicados se aplican al mismo individuo, entonces se infiere un condicional derrotable. De esa manera, se puede relacionar el índice de importancia epistémica del condicional inferido por inducción con el número n de casos en el antecedente de la inferencia. Por consideraciones de espacio no podemos considerar extensivamente este criterio, pero podemos mencionar que como los índices de importancia epistémica conforman un reticulado, es posible ordenarlos en niveles de acuerdo a la distancia del camino más corto hacia el elemento supremo $\mathcal{E}_\top$. De esa manera, es posible calcular el nivel i del índice del condicional inferido en función de la cantidad n de casos regulares.

Tal criterio permitiría por ejemplo inferir, en un contexto donde existen n individuos t_i que son aves y que son voladores, ambas reglas $ave(X) \succ\!\!- vuela(X)$ y $vuela(X) \succ\!\!- ave(X)$. Si exigimos (tal como parece sugerir el esquema de inferencia inductiva 2) que exista por lo menos un caso en el cual la regularidad no se cumpla, entonces se podría inferir la primera regla si existe por lo menos un individuo volador que no sea ave (un insecto, por ejemplo), y/o la segunda si existe un ave no voladora (un pingüino, por ejemplo).

EJEMPLO **4.3** *Consideramos nuestro conocimiento acerca de la religión y los hábitos alimenticios de algunos de nuestros amigos. En particular, aquellos que son buddhistas (b) y vegetarianos (v):*

$E = \{b(john), b(ana), v(john), v(ana), v(mario)\}$.

Por otra parte, es un hecho aceptado que el buddhismo es la religión mayoritaria en la India:

$\mathcal{G} = \{i(X) \succ\!\!\!— b(X)\}$.

¿Que podemos concluir acerca de nuestro nuevo amigo $trilok$, nacido en la India? Realizando una inferencia inductiva en E podremos considerar el condicional $b(X) \succ\!\!\!— v(X)$, el cual, encadenado con $i(X) \succ\!\!\!— b(X)$ nos permite concluir consistentemente $v(trilok)$.

4.3 Abducción

Al igual que la inducción, la abducción no infiere una conclusión, sino más bien infiere la base necesaria para arribar a una conclusión que puede o no observarse. Por lo tanto, también enfrentamos el problema de asignar una importancia epistémica a la sentencia inferida. Pero además hay un nuevo problema, dado que pueden existir varias "explicaciones" inferidas por abducción, de las cuales deberíamos poder elegir aquella que parezca más adecuada.

EJEMPLO **4.4** *En un sistema de diagnóstico ocupacional encontramos que normalmente quien tiene un trabajo (t) normalmente percibe un ingreso (i), normalmente no estudia (e), y paga aporte jubilatorio (j). Quienes estudian normalmente trabajan, y quienes estudian y ganaron una beca (b) perciben un ingreso pero no pagan aporte jubilatorio:*

$$\mathcal{G} = \{ \begin{array}{l} t(X) \succ\!\!\!—_{1} i(X) \\ t(X) \succ\!\!\!—_{2} j(X) \\ t(X) \succ\!\!\!—_{3} \neg e(X) \\ e(X) \succ\!\!\!—_{4} t(X) \\ b(X) \succ\!\!\!—_{5} e(X) \\ b(X) \succ\!\!\!—_{6} i(X)\} \end{array}$$

.

Con respecto a trabajadores, los condicionales δ_2 y δ_3 son más importantes que δ_1, es decir, es más "normal" afirmar que quien trabaja percibe un ingreso y paga su aporte que afirmar que quien trabaja también estudia. Con respecto a estudiantes, los condicionales δ_5 y δ_6 son más importantes que δ_4, es decir, es más normal afirmar que los becados estudian y perciben un ingreso que afirmar que quienes estudian también trabajan. Por último, es posible afirmar que es más normal que quien trabaja no estudia que afirmar que quien estudia también trabaja, es decir, δ_4 es más importante que δ_1:

$\{\delta_4 \prec \delta_5, \delta_4 \prec \delta_6, \delta_1 \prec \delta_2, \delta_1 \prec \delta_3, \delta_4 \prec \delta_1\}$.

En este estado de cosas ¿Qué se puede afirmar de $juan$, de quién sabemos que paga su aporte jubiliatorio?.

Por inferencia abductiva, de $j(juan)$ se sigue $t(juan)$ sin contradicción, y dado que el condicional 1 domina al 4, se sigue tambien $\neg e(juan)$. Por último, aplicando abducción nuevamente, se sigue $\neg b(juan)$.

¿Qué podemos afirmar de ana, de quién sabemos que percibe un ingreso?

En estas condiciones, es posible encontrar abductivamente una justificación en $t(ana)$, y otra en $b(ana)$. Con respecto la primera justificación, se sigue además que $j(ana)$ y que $\neg e(ana)$ y por consiguiente $\neg b(ana)$, es decir, ana percibe un ingreso por trabajar y paga su jubiliación, pero no estudia y por consiguiente no está becada. De la segunda justificación se sigue $e(ana)$, y además $t(ana)$ y $j(ana)$, es decir, ana está becada y por consiguiente estudia, pero además ana trabaja y por consiguiente paga su jubiliación. Como nuestra estructura epistémica es escéptica, la justificación abductiva será aquella que pertenezca a todas las interpretaciones posibles (o permanecer indeterminado si no hay justificación común). Es decir, se acepta que ana trabaja como explicación de su ingreso, y además se predice que ana debe normalmente pagar una jubilación, hecho que habrá que corroborar.

¿Qué podemos afirmar de pedro, de quién sabemos que percibe un ingreso pero que no paga aporte jubilatorio?

En este caso, las justificaciones del ingreso de pedro son idénticas a las justificaciones del ingreso de ana, pero el hecho adicional de que pedro no paga jubilación bloquea la primera, y por lo tanto se concluye que pedro es un estudiante becado.

5 Aplicaciones en el razonamiento científico

En esta sección presentamos una breve reexposición de los resultados principales presentados en [6], para presentar nuevos ejemplos de cómo la arquitectura epistémica desarrollada en este trabajo permite una representación adecuada del conocimiento científico y sus patrones de inferencia asociados. Una forma de describir el propósito de las teorías científicas es la de encontrar el menor conjunto de conocimiento que produzca un *cubrimiento* del conjunto de evidencia E que se pretende sistematizar. Pero este cubrimiento se produce a través de un conjunto de procedimientos de inferencia. Hempel [11] fue el primero en proponer que la evidencia debe inferirse de las leyes, y no a la inversa, la cual era la visión epistemológica tradicional. 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*. Hempel propone que la lógica de la predicción y de la explicación proceden según un mismo esquema $\mathcal{L} \vdash e$, donde $\mathcal{L}$, 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 es denominado *contexto de predicción* y *contexto de explicación*, respectivamente. La sistematización por medio de este esquema se denominó *paradigma hipotético-deductivo*, dado que el *explanans* constituye una pieza de conocimiento hipotético, de la cual se debe deducir la evidencia.

El procedimiento de inferir el *explanans* no puede ser deductivo, es decir, $\mathcal{L}$ nunca puede ser *verdadera*. Una conclusión, señalada por Popper [26] 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- sirven para mostrar que una teoría es falsa.

Este comportamiento muestra que el esquema hipotético-deductivo es pragmáticamente poco adecuado y que no describe adecuadamente el comportamiento real, dado que mientras una teoría produzca resultados positivos no será completamente abandonada. Este hecho, observado por Lakatos [13], fue el inspirador de su reconstrucción de la dinámica de las teorías

científicas, denominada por él *programas de investigación*. Lakatos elaboró una reconstrucción de la dinámica de las teorías científicas teniendo en cuenta el contexto histórico que tiene la dinámica de una teoría científica. Este aspecto determina que el *método* científico no sea único, es decir, distintas comunidades adoptan distintas prácticas. Por lo tanto, en una misma ciencia pueden coexistir distintas teorías para explicar un mismo fenómeno, cada una propugnada por una parte de la comunidad científica que adhiere a un determinado aspecto metodológico. Cada una de estas teorías, junto con su metodología subyacente, es un *programa de investigación* que compite con los demás. Los programas de investigación son estructuras que incluyen a las teorías científicas, integrándolas con un conjunto de procedimientos de inferencia y un conjunto periférico de hipótesis auxiliares.

El *núcleo* de un programa es un conjunto de conocimiento que se considera central, y que define al programa 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 falsación. Esta protección consiste básicamente en implementar un *cinturón protector* (en la terminología original) de hipótesis auxiliares, que impiden que el núcleo sea falsado. 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 h particular al caso e tal que de la teoría aumentada con h se siga e. Si dicha hipótesis no es compatible con el resto de la teoría, entonces algo en la misma deberá corregirse. 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. Como vimos, existen los principios generales (expresados en la sentencia P a continuación), las leyes L que se siguen lógicamente de los principios, pero que pragmáticamente tienen un uso más adecuado, y la evidencia, es decir, los fenómenos directamente constatables.

EJEMPLO **5.1** *Nuestro conocimiento acerca de gravitación se reduce a:*

- $\mathcal{K}$: *Existe una fuerza que atrae a los objetos (masivos) entre sí.*
  $\forall X, Y. o(X) \wedge o(Y) \Rightarrow a(X, Y)$
- $\mathcal{L}$: *Los objetos son atraídos hacia la tierra.*
  $\forall X. o(X) \Rightarrow a(tierra, X)$
- e_1: *Esta piedra es atraída hacia la tierra.*
  $a(tierra, p)$
- e_2: *Este globo de hidrógeno no es atraído hacia la tierra.*
  $\neg a(tierra, g)$

La primer sentencia es un principio, la segunda es una ley, y las dos últimas son evidencia. La ley se deduce como caso particular del principio, y es por lo tanto utilizada por su mayor valor pragmático. Sin embargo, si bien la ley permite explicar (o predecir) la tercer sentencia, fracasa con la cuarta. Es decir, tenemos $\mathcal{K} \vdash \mathcal{L}$, $\mathcal{L} \vdash e_1$, *pero* $\mathcal{L} \nvdash e_2$. *Esto, sin embargo, no lleva a abandonar la ley, 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* $L, C \vdash E$, *donde* E *cubre tanto a* e_1 *como a* e_2, *y* C *expresa las condiciones particulares a cada objeto.*

Los resultados de este trabajo se aplican para representar el conocimiento de una mejor manera que con la lógica clásica. Para sistematizar un conjunto de evidencia se modifica el *explanans* Hempeliano para incluír un conjunto de conocimiento plausible y un conjunto condicionales derrotables. De ese modo tenemos que $\{\mathcal{K} \cup \mathcal{G} \cup P\}$ *justifica* e. En este caso, $\mathcal{K} \cup \mathcal{G}$ reemplaza al núcleo $\mathcal{L}$ Hempeliano, y P al conjunto C de condiciones particulares.

EJEMPLO **5.2** *(Ejemplo 5.1 revisitado).*

$\mathcal{K}$: *Existe una fuerza que atrae a los objetos masivos entre sí.*
$\forall X, Y.o(X) \wedge o(Y) \Rightarrow a(X, Y)$
$\mathcal{G}$: *Los objetos tienden a caer hacia la tierra.*
$o(X) \succ\!\!- a(tierra, X)$
e_1: *Esta piedra cae hacia la tierra.*
$a(tierra, p)$
e_2: *Este globo de hidrógeno no cae hacia la tierra.*
$\neg a(tierra, g)$

En este ejemplo, $\mathcal{G}$, por estar sujeta a excepciones de origen desconocido, se utiliza como regla tentativa. De ese modo e_1 está justificada dentro de la teoría, mientras que e_2 constituye una excepción. Más adelante, una amiga que ha leído a Arquímedes y a Torricelli nos sugiere la presencia de una atmósfera como causa de algunas de las excepciones en nuestra teoría:

$\mathcal{K}$: *Existe una fuerza que atrae a los objetos masivos entre sí.*
$\forall X, Y.o(X) \wedge o(Y) \Rightarrow a(X, Y)$
$\mathcal{G}_1$: *Los objetos más pesados que el aire tienden a caer hacia la tierra.*
$o(X) \wedge p(X) \succ\!\!- a(tierra, X)$
$\mathcal{G}_2$: *Los objetos menos pesados que el aire no tienden a caer hacia la tierra.*
$o(X) \wedge \neg p(X) \succ\!\!- \neg a(tierra, X)$
p_1: *Esta piedra es más pesada que el aire.*
$p(p)$
p_2: *Este globo de hidrógeno es menos pesado que el aire.*
$\neg p(g)$
e_1: *Esta piedra cae hacia la tierra.*
$a(tierra, p)$
e_2: *Este globo de hidrógeno no cae hacia la tierra.*
$\neg a(tierra, g)$

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

En el ejemplo anterior, utilizamos implicitamente la teoría escéptica para encontrar las conclusiones buscadas. El paso siguiente consiste en encontrar un modelo de la búsqueda y la revisión de justificaciones alternativas dentro del contexto de descubrimiento por medio de los patrones de inferencia ampliativa. Dicho contexto está caracterizado por la generación de hipótesi s condicionales, las cuales –si están confirmadas por la experiencia– prestan apoyo a la teoría. Es decir, se generan conjeturas con bases firmes, seguido de pruebas rigurosas. El proceso de conjetura normalmente es asistemático, interviendo la intuición e imaginación

dentro de los límites de la disciplina. Habitualmente se espera que las hipótesis de trabajo en el contexto de descubrimiento respondan a criterios como contrastabilidad, repetibilidad, simplicidad y coherencia. Una hipótesis que genere experimentos que se confirman en algunos casos y fracasan en otros no es en principio desechada.

Estas características hacen del razonamiento científico un campo de muy dificil sistematización y formalización, dado que el patrón de razonamiento va de lo particular a lo general (lo opuesto a la deducción). La forma de expresar las teorías científicas expresada en este trabajo, junto con la importancia epistémica asignada a cada parte del conocimiento de las mismas, permite avanzar hacia el diseño de experimentos hipotéticos y su subsecuente análisis. Analizaremos primero la forma de tratar el razonamiento hipotético, para luego establecer una manera de generar hipótesis por inducción y contrastarlas por medio de dicho razonamiento.

La manera de implementar el razonamiento hipotético en este marco consiste en tratar a la premisa hipotética (problemática o contrafáctica) como un conocimiento particular de máxima plausibilidad. Por lo tanto el conflicto que se produce al introducirla se establece a partir de otras piezas de conocimiento particular. Podemos citar tres casos de conflicto. En el primero, la premisa hipotética es contradictoria con la consecuencia de un razonamiento *default* basado en conocimiento particular (el cual puede ser plausible).

EJEMPLO **5.3** *Conocemos* pingüino(*opus*) *y* pingüino$(X) \mathrel{>\!\!\!\sim} \neg vuela(X)$*, y nos planteamos la hipótesis* "¿Qué sucedería si *opus* volara?"*. En este caso la conclusión que se prefiere es que* "Si *opus* volara entonces es una excepción a la regla de que normalmente los pingüinos no vuelan"*. Esto es así porque si hay una razón plausible para aceptar* pingüino(*opus*)*, entonces a dicha razón se le otorga mayor importancia epistémica que a la regla* default.

Los demás casos de razonamiento hipotético (cuando la premisa hipotética es contradictoria con un conjunto de conocimiento particular o cuando la premisa hipotética es contradictoria con la consecuencia de un razonamiento deductivo basado en conocimiento particular) son también manejados adecuadamente. Estos ejemplos abarcan las posibilidades de razonamiento hipotético para todos los tipos de conocimiento presentados. Otro 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 con condicionales [40] como por ejemplo el test de Ramsey, y el cambio de teorías [1], pueden 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 $\mathcal{E}_\top$ 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.

Por último, ilustraremos algunos de los procedimientos de inferencia típicos en el desarrollo de una teoría científica a través de ejemplos idealizados. Consideraciones de espacio no nos permiten trabajar con ejemplos reales de la historia del pensamiento científico, pero es posible ver que el procedimiento abstracto seguido en cada uno de ellos es similar.

EJEMPLO **5.4** *Sea la teoría* $\mathcal{T} = \{a, a \mathrel{>\!\!\!\sim} b\}$*. Esta teoría predice* b*. ¿Qué sucede si* b *no se observa, es decir, si hay evidencia cierta que* $\neg b$*? En esta situación se pueden dar varios casos.*

En el primero, se crea una teoría $\mathcal{T}_1 = \{a, \neg b, a \mathrel{>\!\!\!\sim} b\}$*, con* $\{a \prec \neg b, a \prec (a \mathrel{>\!\!\!\sim} b)\}$*. Según* $\mathcal{T}_1$*, la causa del fracaso es debida a* a*, que no está debidamente justificada, pero* $a \mathrel{>\!\!\!\sim} b$ *se puede seguir manteniendo, es más, crea la presuposición de que* $\neg a$ *que habrá que corroborar.*

En el segundo caso, se crea otra teoría $\mathcal{T}_2 = \{a, \neg b, a \succ\!\!\!- b\}$, *con* $\{a \prec \neg b, (a \succ\!\!\!- b) \prec a\}$. *Según* $\mathcal{T}_2$, *la causa del fracaso es debida a* a, $a \succ\!\!\!- b$ *que está falsada por la evidencia, pero el dato* a *se puede seguir manteniendo.*

Pueden existir casos en los que se agregan hipótesis auxiliares c *para proteger a la "ley" de ser falsada, creándose una teoría* $\mathcal{T}_3 = \{a, c, \neg b, a \succ\!\!\!- b\}$, *con*$\{(a, c) \succ\!\!\!- \neg b\}$. *Según* $\mathcal{T}_3$, *la causa del fracaso es debida a que* $a \succ\!\!\!- b$ *sistematiza solo una parte del conocimiento, pero debe existir otra ley* $(a, c) \succ\!\!\!- \neg b$ *que la completa para la situación particular* c.

En este último ejemplo podemos ver la evolución formal 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 $\mathcal{T}$ tradicionalmente aceptada, la cual fracasaba en algunos casos particulares. La teoría, lejos de ser abandonada, fue protegida o bien rechazando los datos y buscando nuevos (caso $\mathcal{T}_1$), o bien, cuando estos datos se corroboraban (algo que llevaría al caso $\mathcal{T}_2$), buscando condiciones particulares y nuevas leyes que la completaran (caso $\mathcal{T}_3$). Este modo de proceder fue lo que Lakatos denominó *heurística negativa*.

6 Conclusiones y trabajo futuro

Se presentó una interpretación plausibilístca de los patrones ampliativos de inferencia. La misma se fundamenta en un sistema de representación de conocimiento y razonamiento basado en la inferencia exductiva a partir de un orden de importancia epistémica. El mismo permite representar conocimiento tentativo de distintos orígenes (información plausible y condicionales derrotables), y es adecuado para incorporar distintos patrones de inferencia ampliativos, como el razonamiento no monotónico, la inducción y la abducción. Se propuso una semántica escéptica para caracterizar el conjunto de conclusiones, así como un procedimiento efectivo de prueba correcto y completo. Por último, se presentaron algunas correspondencias importantes entre este sistema y los modelos de razonamiento científico.

Actualmente estamos trabajando en mejorar el modelo de razonamiento científico basándonos en el paradigma hipotético-deductivo de Hempel y en los programas de investigación de Lakatos. Un punto importante, en particular, es representar la *competencia* entre distintos programas de investigación. Esto trata de reflejar una realidad histórica, en la cual varias teorías científicas coexisten en el tiempo como posibles explicaciones de ciertos fenómenos, hasta que son completamente abandonadas.

Agradecimiento: Algunas de las ideas aquí presentadas fueron discutidas con Fernando Tohmé, Guillermo Simari, Juan Manuel Torres y Jorge Roetti. Este trabajo fue parcialmente financiado con un subsidio de la Secretaría de Ciencia y Técnica de la Universidad Nacional del Sur.

Referencias

[1] Carlos Alchourón, Peter Gärdenfors, y David Makinson. On The Logic of Theory Change. *The Journal of Symbolic Logic*, 50(2):510–530, 1985.

[2] Dana Angluin y Carl Smith. Inductive Inference: Theory and Methods. *ACM Computer Surveys*, 15(3):237–269, 1983.

[3] Craig Boutilier y Verónica Becher. Abduction as Belief Revision. *Artificial Intelligence*, 77(1):43–94, 1995.

[4] Roderick M. Chisholm. *Theory of Knowledge*. Prentice Hall, Englewood Cliffs, New Jersey, 1977.

[5] Keith L. Clark. Negation as Failure. En Herve Gallaire y Jack Minker, editores, *Logic and Data Bases*, pages 293–322. Plenum Press, New York, 1978.

[6] 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.

[7] Claudio Delrieux y Guillermo Simari. Formalizing Plausible Reasoning. En *Proceedings of the XV International Conference of the Chilean Computer Society*, páginas 147–158, Arica, Chile, 1995. SCCC.

[8] David W. Etherington. *Reasoning with Incomplete Information*. Morgan Kaufmann Publishers, Los Altos, CA, 1988.

[9] Héctor Geffner y Judea Pearl. Conditional Entailment: Bridging Two Approaches to Default Reasoning. *Artificial Intelligence*, 53(2-3):209–244, 1992.

[10] Carl G. Hempel. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press, New York, 1965.

[11] Carl G. Hempel y Paul Oppenheim. The Logic of Explanation. *Philosophy of Science*, 15:135–175, 1948.

[12] Kurt Konolige. Abduction versus Closure in Causal Theories. *Artificial Intelligence*, 53(2-3):255–272, 1992.

[13] Imre Lakatos. *Proofs and Refutations. The Logic of Mathematical Discovery*. Cambridge University Press, 1976.

[14] Jinxin Lin. A Semantics for Reasoning Consistently in the Presence of Inconsistency. *Artificial Intelligence*, 86(1):75–95, 1996.

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

[16] Eliezer L. Lozinskii. Resolving Contradictions: A Plausible Semantics for Inconsistent Systems. *Journal of Automated Reasoning*, 12(1):1–31, 1994.

[17] John McCarthy. Circumscription — A Form of Non-monotonic Reasoning. *Artificial Intelligence*, 13(1-2):27–39,171–172, 1980.

[18] John McCarthy. Applications of Circumscription to Formalizing Common-Sense Knowledge. *Artificial Intelligence*, 28(1):89–116, 1986.

[19] Drew McDermott y Jon Doyle. Non-monotonic Logic I. *Artificial Intelligence*, 13(1,2):41–72, 1980.

[20] Robert C. Moore. Semantical Considerations of Nonmonotonic Logic. *Artificial Intelligence*, 25(1):75–94, 1985.

[21] Donald Nute. Defeasible Reasoning. En James H. Fetzer, editor, *Aspects of Artificial Intelligence*, páginas 251–288. Kluwer Academic Publishers, Norwell, MA, 1988.

[22] Tarcisio Pequeno. A Logic for Inconsistent Nonmonotonic Reasoning. Technical Report 90-6, Department of Computing, Imperial Logic, London, 1990.

[23] Gadi Pinkas y Ronald Loui. Reasoning from Inconsistency: A Taxonomy of Principles for Resolving Conflict. En *Proceedings of the Third Knowledge Representation and Reasoning Conference*, páginas 453–460, Los Altos, CA, 1991. Morgan Kaufmann Publishers.

[24] David Poole. A Methodology for Using a Default and Abductive Reasoning System. Technical Report DCS-UW, University of Waterloo, 1988.

[25] David L. Poole. On the Comparison of Theories: Preferring the Most Specific Explanation. En *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, páginas 144–147, Los Altos, CA, 1985. International Joint Conference on Artificial Intelligence, Morgan Kaufmann Publishers.

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

[27] Karl Popper. *Conjectures and Refutations.* Routledge and Kegan Paul, London, 1963.

[28] Teodor Przymusinski. On the Relationship Between Logic Programming and Non-monotonic Reasoning. En *Proceedings of the Sixth National Conference on Artificial Intelligence*, páginas 444–448, Los Altos, CA, 1988. American Association for Artificial Intelligence, Morgan Kaufmann Publishers.

[29] Terry Rankin. Could Nonmonotonic Inference Ever Be Deductively Valid. Technical Report ACMC-01-0009, Advanced Computational Methods Center, University of Georgia, December 1985.

[30] Raymond Reiter. On Closed World Data Bases. En Herve Gallaire y Jack Minker, editores, *Logic and Data Bases*, páginas 55–76. Plenum Press, New York, 1978.

[31] Raymond Reiter. A Logic for Default Reasoning. *Artificial Intelligence*, 13(1,2):81–132, 1980.

[32] Raymond Reiter. A Theory of Diagnosis from First Principles. *Artificial Intelligence*, 32:57–95, 1987.

[33] Nicholas Rescher. *Plausible Reasoning.* Van Gorcum, Dodrecht, 1976.

[34] Nicholas Rescher. *Dialectics, a Controversy-Oriented Approach to the Theory of Knowledge.* State University of New York Press, Albany, 1977.

[35] John Reynolds. Transformational Systems and the Algebraic Structure of Atomic Formulas. *Machine Intelligence*, 5:135–151, 1970.

[36] Nico Roos. A Logic for Reasoning with Inconsistent Knowledge. *Artificial Intelligence*, 57(1):69–104, 1992.

[37] J. C. Shepherdson. Negation in Logic Programming. En Jack Minker, editor, *Foundations of Deductive Databases and Logic Programming*, páginas 19–88. Morgan Kaufmann, Los Altos, California, 1987.

[38] Paul Thagard. *Computational Philosophy of Science.* MIT Press, Cambridge, MA, 1988.

[39] David S. Touretzky. *The Mathematics of Inheritance Systems.* Morgan Kaufmann Publishers, Los Altos, CA, 1986.

[40] Johan van Benthem. *Intensional Logics.* CSLI/SRI International, Stanford, second edition, 1988.

[41] Gerd Wagner. Ex Contradictione Nihil Sequitur. En *Proceedings of the Twelvth International Joint Conference on Artificial Intelligence*, páginas 538–543, Los Altos, CA, 1993. International Joint Conference on Artificial Intelligence, Morgan Kaufmann Publishers.