

Modelos Computacionales del Razonamiento en las Teorías Científicas

Claudio Delrieux
Universidad Nacional del Sur, Bahía Blanca - ARGENTINA
claudio@acm.org

Resumen

Las teorías científicas pueden considerarse como extensiones de las teorías lógicas, en las que se incorpora un conjunto heterogeneo y posiblemente inconsistente de conocimiento tentativo (leyes empíricas, conjeturas, hipótesis auxiliares, etc.). Dicho conocimiento está organizado por medio de una jerarquía que representa una dimensión pragmática que tiene en cuenta los posibles beneficios en el uso de dicho conocimiento por medio de patrones de inferencia, englobados dentro de lo que usualmente se denomina método científico. En este trabajo se propone formalizar algunos aspectos de la dinámica de las teorías científicas, tomando como base la propuesta de Lakatos denominada Programas de Investigación Científica. En particular, consideramos que un programa es representable por medio de teorías no monotónicas. También se analiza el efecto en los programas de la aparición de nueva evidencia, y se propone un mecanismo de comparación de programas en función del progreso empírico o del grado de confirmación de los mismos.

PALABRAS CLAVE: Inteligencia Artificial — Representación del Conocimiento y Razonamiento — Teoría de la Ciencia

1 Introducción

La ciencia en desarrollo se desenvuelve por medio de la elaboración más o menos cotidiana de *teorías* por parte de los científicos. Éstas, 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 [9, 14, 15]. 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*. 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. Por lo tanto, la representación del conocimiento científico y de sus patrones de inferencia asociados es un objetivo que plantea una serie de dificultades en el marco de la implementación de sistemas de inteligencia artificial [2].

En este trabajo se propone una formalización de diversos aspectos de la dinámica de las teorías científicas, tomando como base la propuesta de Lakatos denominada *Programas de Investigación Científica* [12]. Según ésta, en una disciplina dada existen diversos programas de investigación que compiten entre sí en función de los progresos que realizan. Un programa consiste en un núcleo (o teoría) y un cinturón protector de hipótesis auxiliares. La metodología de investigación en estos programas consiste en proponer procedimientos de conjetura y contrastación para procurar el avance del programa, y de procedimientos heurísticos para defenderlo de las refutaciones. En este contexto, dos o más programas o “teorías” (subconjuntos del conocimiento disponible) progresan de una manera competitiva, buscando adecuadas predicciones o explicaciones referidas a su dominio. Cada programa es llevado adelante por un grupo de investigadores, aquellos que por la preferencia asignada al conjunto de conocimiento existente coinciden en aceptar como definitivo un grupo de hipótesis “centrales”.

Pero los programas o teorías no son estáticos, dado que existe un proceso de adopción o rechazo de elementos de conocimiento “periféricos” (resultados experimentales, conjeturas, etc.) que no afecten al “núcleo” de la teoría, que la define como tal. El éxito de una teoría frente a sus competidoras puede eventualmente llevarla a su establecimiento hegemónico, es decir, a que todos los grupos de investigación abandonen a los programas competidores. El sistema presentado en este trabajo permite representar adecuadamente la estructura de un programa de investigación de acuerdo a esta formulación, se considera el efecto en los programas de la aparición de nueva evidencia, y se propone un mecanismo de comparación de programas en función del progreso empírico o del grado de confirmación de los mismos.

2 Aspectos formales en la Teoría de la Ciencia

En esta Sección veremos cómo es posible formalizar algunas de las ideas imperantes en la epistemología y el razonamiento científico utilizando un lenguaje lógico. Este tema está siendo cada vez más el centro de atención en la investigación actual en Inteligencia Artificial [6, 7] Dentro del método 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 uno de los principales objetos de estudio de la teoría de la ciencia. El conocimiento científico se ordena y configura en estructuras complejas. Las unidades

de organización más destacadas dentro de esta estructura 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 justificado y el método científico fuese deductivamente válido, entonces no habría diferencia entre ambos tipos de teorías. Sin embargo, las teorías científicas involucran tipos de conocimiento cuya justificación es problemática y mecanismos de inferencia ampliativos. Por lo tanto estas teorías no tienen el *status* de ser deductivamente válidas.

Podemos describir por lo menos tres dominios o niveles de enunciados dentro de una teoría [1, 10]. Dado un conjunto de fenómenos, entidades o propiedades de un determinado aspecto de la realidad del cual una ciencia se ocupa, el primer nivel \mathcal{N}_1 es un conjunto de enunciados particulares que representa los diversos estados de cosas posibles en dicho dominio. Este nivel es esencialmente *proposicional* (en el sentido que Levesque [13] le da al término) dado que un enunciado cualquiera $p(a)$ se interpreta como “*es un hecho empíricamente observable que el objeto (entidad, fenómeno) a tiene la propiedad (característica, circunstancia) p*”. Lo que comunmente denominamos *evidencia* está dentro de este nivel. Por ejemplo, “*es un hecho empíricamente observado que Opus es un pingüino*” se representa con $pingüino(opus)$. 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. Estos enunciados se obtienen por observación o experimentación, y se justifican por la experiencia.

El segundo nivel \mathcal{N}_2 está constituido por generalizaciones empíricas o accidentales [8, 14]. 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 (característica, circunstancia) observable p, normalmente tienen la propiedad q*”. Para representar estos enunciados en un lenguaje lógico, debemos extender al mismo con un “condicional derrotable” o “implicación *prima facie*” $p(X) \succ q(X)$. Este tipo de enunciados condicionales, como vimos más arriba, se pueden obtener por abstracción (clases de equivalencia entre enunciados del nivel anterior), o por analogía (isomorfismos con otras estructuras de enunciados). Como tales, su inferencia depende de la imaginación y la creatividad, y su justificación es enormemente problemática. Normalmente estos enunciados se aceptan gradualmente por la comunidad, pero una vez establecidos, la misma comunidad que se comportó escepticamente, ahora está comprometida a utilizarlos y retenerlos aún al gran costo de rechazar evidencia [15].

El tercer nivel \mathcal{N}_3 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 también *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. Este nivel permite explicar, predecir, descubrir y sistematizar el conocimiento de un determinado aspecto de la realidad. En este nivel también podemos contar los enunciados o *principios puente*, que establecen una correspondencia entre términos teóricos y observables. Los enunciados en este nivel normalmente son leyes universalmente cuantificadas, dado que expresan un conocimiento analítico que se justifica por provenir de primeros principios o de las definiciones. Un ejemplo proviene de las clasificaciones. El término *ave*, por ejemplo, es un prototipo que se establece en una clasificación zoológica. De esa manera, el enunciado “*Todos los pingüinos son aves*” es verdadero por definición, y se representa como $\forall X.(ave(X) \Rightarrow pingüino(X))$. Puede darse también que un principio interno exprese conocimiento sintético. Por ejemplo, las leyes de la física establecen que existe una fuerza que atrae a los objetos masivos entre sí lo cual se representa como $\forall X, Y.(objeto(X) \wedge objeto(Y) \Rightarrow atracc(X, Y))$.

2.1 El paradigma hipotético-deductivo

Las teorías forman parte de una disciplina, pero no las sistematizan exhaustivamente. El propósito de las teorías es en principio económico: se busca 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. La primera formalización importante de una teoría de la ciencia, como vimos más arriba, fue propuesta en el Círculo de Viena, y se basó fundamentalmente en el uso de procedimientos inductivos. El esquema subyacente consistía en mostrar que las leyes científicas se infieren de la evidencia. Este acercamiento fue encontrando dificultades insalvables.

Hempel [9] fue el primero en proponer que la evidencia debe inferirse de las leyes, y no a la inversa. Hempel, propone que la lógica de la predicción y de la explicación proceden según un mismo esquema $\mathcal{T} \vdash e$, donde \mathcal{T} , el *explanans* es un conjunto de leyes, y $e \in E$, el *explanandum* es el fenómeno o hecho a explicar. 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 única diferencia constituye el *contexto* dentro del cual se utiliza el esquema, el cual es denominado *contexto de predicción* y *contexto de explicación*, respectivamente. Es conveniente notar que en este esquema, el *explanandum* pertenece al primer dominio o nivel ($e \in \mathcal{N}_1$), mientras que el *explanans* pertenece a los otros dos ($\mathcal{T} \subseteq \mathcal{N}_2 \cup \mathcal{N}_3$).

La sistematización por medio de este esquema se denominó *paradigma hipotético-deductivo* o deductivo-nomológico, dado que el *explanans* está constituido por conocimiento hipotético, normalmente incluyendo como piezas centrales algunas “leyes”, y del cual se debe deducir la evidencia. El procedimiento de inferir el *explanans* no puede ser deductivo, es decir, \mathcal{T} nunca puede ser *verdadera*. Una conclusión, señalada por

Popper 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, fue el inspirador de su reconstrucción de la dinámica de las teorías científicas, denominadas por él *programas de investigación*, la cual formalizaremos más abajo.

2.2 Los programas de investigación científica de Lakatos

El nombre de *Programas de Investigación Científica* fue propuesto por Lakatos [11] para describir su reconstrucción de la dinámica de las teorías científicas. Lakatos distingue la *historia interna* y 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 no sea único, es decir, distintas comunidades o grupos adoptan distintas prácticas o siguen distintas estrategias. 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 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. 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 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 a la teoría como tal. 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* (en la terminología original) 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*. 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).

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. El éxito de una teoría consiste en predecir o explicar nuevas observaciones, es decir, que la teoría produzca un *progreso empírico* en su dominio. Por lo tanto, en Lakatos una teoría sobrevive a las anomalías en tanto siga siendo progresiva. De esa forma, el *programa de investigación* queda determinado por su núcleo teórico (el que en definitiva le da identidad como tal), y el *cinturón protector* de hipótesis auxiliares, cuyo objetivo exclusivo es proteger al núcleo de la refutación. Dado que, entonces, los programas se defienden de las anomalías generando hipótesis auxiliares, ¿cuándo y por qué debería ser abandonado un programa? Lakatos da para esto una respuesta más cercana a la pragmática Kuhniana que al realismo Popperiano: mientras el programa siga siendo progresivo, nunca será definitivamente abandonado. La única situación en la que esto puede no ser así es que un programa quede completamente subsumido dentro de otro, el cual tenga menor cantidad de hipótesis *ad hoc*. Pero esto también es, si se quiere, una cuestión de opiniones. En cambio, un programa cuya única actividad consiste en defenderse de las refutaciones, sin generar progreso, se denomina un programa *degenerado*, el cual, de existir opciones, normalmente es abandonado.

3 Formalizando los *programas* de Lakatos

En esta Sección propondremos una formalización de los *Programas*, en términos de los sistemas de razonamiento no monotónicos. También esbozaremos algunas ideas tendientes a comparar entre sí diferentes *programas* en función de algunas características ventajosas, como ser el progreso empírico y el éxito relativo, dejando planteadas otras posibilidades como por ejemplo el poder sistematizador, o la menor cantidad de hipótesis *ad hoc*. Esta formalización presupone que el conocimiento válido \mathcal{K} es conocido y universalmente aceptado por todos los programas, al igual que la evidencia E . Utilizaremos como mecanismo de inferencia subyacente a un sistema abstracto de razonamiento argumentativo [16, 17]. Es decir, las predicciones o explicaciones en los programas se obtienen construyendo argumentos.

DEFINICIÓN 1 Dado un conjunto \mathcal{K} de conocimiento válido y un conjunto E de evidencia (literales de base), un programa de investigación $P = \langle \mathcal{T}, C \rangle$ consta de una teoría no monotónica \mathcal{T} y un cinturón protector C . La teoría \mathcal{T} está representada por medio de un conjunto de condicionales derrotables $a(X) \succ\!\!\!-\ b(X)$, y C por medio de literales de base que representan las hipótesis *ad hoc*. Un programa de investigación $P = \langle \mathcal{T}, C \rangle$ predice (o explica) a un literal de base e cuando existe un argumento para e a partir de $\mathcal{K}, \mathcal{T}, C$ y E .

Es importante recordar que la existencia de un argumento para un literal no es condición suficiente para asumir que la teoría argumentativa acepta a dicho literal como conclusión, dado que pueden existir argumentos en contra que lo derroten. Esto

es análogo a lo que ocurre en un programa de investigación, donde el programa puede realizar predicciones para un observable, pero también predicciones contradictorias con el mismo observable. Por lo tanto deberíamos repasar cuáles son las “actitudes proposicionales” que la teoría argumentativa nos permite asumir respecto del literal e .

DEFINICIÓN 2 Dado un programa P y un literal de base e , se pueden dar los casos:

	<i>Definitiva</i>	<i>Revisable</i>
<i>Aceptar e</i>	<i>Argumentos solamente para e</i>	<i>Argumentos tanto para e como para $\neg e$ pero existe un argumento no derrotado para e</i>
<i>Aceptar $\neg e$</i>	<i>Argumentos solamente para $\neg e$</i>	<i>Argumentos tanto para e como para $\neg e$ pero existe un argumento no derrotado para $\neg e$</i>
<i>Indeciso</i>	<i>No hay argumentos ni para e ni para $\neg e$</i>	<i>Argumentos tanto para e como para $\neg e$ pero no existen argumentos no derrotados</i>

Dado un programa P y un literal de base e el cual se confirma en la experiencia, se puede dar solamente uno de entre los siguientes casos:

<i>Predicción de P</i>	<i>Status de P luego de la confirmación de e</i>
<i>Aceptación definitiva de e</i>	<i>P se confirma</i>
<i>Aceptación revisable de e</i>	<i>P se confirma parcialmente</i>
<i>Aceptación definitiva de $\neg e$</i>	<i>Anomalía grave</i>
<i>Aceptación revisable de $\neg e$</i>	<i>Anomalía parcial</i>
<i>Indecisión definitiva</i>	<i>Hecho sorprendente</i>
<i>Indecisión revisable</i>	<i>“Laguna”</i>

EJEMPLO 1 Supongamos que nuestro conocimiento, evidencia, y suposiciones acerca de las propiedades de algunos animales es el siguiente:

$$\begin{aligned} \mathcal{K} &= \{pingüino(X) \Rightarrow ave(X)\}. \\ \mathcal{T}_1 &= \{ave(X) \succ\!\!\prec vuela(X), pingüino(X) \succ\!\!\prec \neg vuela(X), en-avión(X) \succ\!\!\prec vuela(X)\}. \\ E &= \{pingüino(Opus), ave(Pintín), ave(Tweety), pingüino(Schmuck), en-avión(Schmuck)\}. \end{aligned}$$

Por el momento, nuestro programa no requiere hipótesis auxiliares $C_1 = \{\}$. En estas condiciones, las predicciones que hace nuestro programa $P_1 = \langle \mathcal{T}_1, C_1 \rangle$ con el conocimiento subyacente \mathcal{K} y la evidencia E se basan en lo siguiente:

1. Existe un argumento para $vuela(Opus)$ (por ser ave), pero hay otro argumento para $\neg vuela(Opus)$ (por ser pingüino) que es más específico y que lo derrota. Por lo tanto P_1 predice $\neg vuela(Opus)$.

2. Existe un argumento para $vuela(Tweety)$, por lo tanto P_1 predice dicho literal.
3. Existe un argumento para $vuela(Pintín)$, por lo tanto P_1 predice dicho literal.
4. Existe un argumento para $vuela(Schmuck)$ (por estar en avión) y otro argumento para $\neg vuela(Schmuck)$ (por ser pingüino). Ambos argumentos se derrotan mutuamente, por lo que P_1 no hace predicciones al respecto.

Supongamos ahora que realizamos observaciones acerca de los individuos voladores, encontrando la siguiente evidencia:

$$E_r = \{\neg vuela(Opus), \neg vuela(Tweety), vuela(Gringa), vuela(Pintín), vuela(Schmuck)\}.$$

En dicho caso, podemos ver que nuestro programa P_1 tiene el siguiente status frente a las nuevas observaciones: con $\neg vuela(Opus)$ tiene una confirmación parcial, con $\neg vuela(Tweety)$ tiene una anomalía grave, con $vuela(Gringa)$ tiene una observación sorprendente, con $vuela(Pintín)$ tiene una confirmación, y con $vuela(Schmuck)$ tiene una laguna.

Por el momento, no estableceremos diferencias entre confirmaciones y confirmaciones parciales, lagunas y hechos sorprendentes, y anomalías parciales o graves. La distinción será importante más adelante, cuando tratemos en la próxima Sección de formalizar algunos aspectos de la dinámica de los programas, es decir, cuándo y cómo el programa debe modificarse para adaptarse a las nuevas observaciones.

De esa forma, denominaremos V al conjunto de casos donde la evidencia confirma total o parcialmente al programa P , I las indeterminaciones o casos sorprendentes o lagunas para P , y F los casos refutatorios o anomalías parciales o graves. En el ejemplo anterior, podemos ver que $V_1 = \{\neg vuela(Opus), vuela(Pintín)\}$, $I_1 = \{vuela(Gringa), vuela(Schmuck)\}$, $F_1 = \{\neg vuela(Tweety)\}$.

DEFINICIÓN 3 Un programa P_1 es empíricamente más progresivo que otro P_2 siempre que tenga estrictamente mayor cantidad de casos confirmatorios. Es decir,

$$P_1 \prec_{prog} P_2 \Leftrightarrow V_2 \subset V_1.$$

P_1 es más exitoso que P_2 si cada caso confirmatorio de P_2 es también confirmatorio de P_1 , cada caso refutatorio de P_1 es también refutatorio de P_2 , pero hay por lo menos un caso confirmatorio de P_1 que no es confirmatorio de P_2 o por lo menos un caso refutatorio de P_2 que no es refutatorio de P_1 . Es decir,

$$P_1 \prec_{ex} P_2 \Leftrightarrow V_2 \subseteq V_1, F_1 \subseteq F_2, \exists v \in V_1 - V_2 \vee \exists f \in F_2 - F_1$$

EJEMPLO 2 Supongamos estar en una situación similar a la del Ejemplo 1. A partir de la evidencia previa E_d del ejemplo, un programa P_2 edificado a partir una segunda teoría \mathcal{T}_2 realizará las siguientes predicciones:

$$\mathcal{T}_2 = \{ave(X) \succ\!\!\prec vuela(X), pingüino(X) \succ\!\!\prec \neg vuela(X)\}.$$

1. Existe un argumento para $vuela(Opus)$ (por ser ave), pero hay otro argumento para $\neg vuela(Opus)$ (por ser pingüino) que es más específico y que lo derrota. Por lo tanto P_2 predice $\neg vuela(Opus)$.
2. Existe un argumento para $vuela(Tweety)$, por lo tanto P_2 predice dicho literal.
3. Existe un argumento para $vuela(Pintín)$, por lo tanto P_2 predice dicho literal.
4. Existe un argumento para $\neg vuela(Schmuck)$ (por ser pingüino), por lo que P_2 predice dicho literal.

Si luego nos confrontamos a las mismas observaciones

$$E_r = \{\neg vuela(Opus), \neg vuela(Tweety), vuela(Gringa), vuela(Pintín), vuela(Schmuck)\},$$

podemos ver que el nuevo programa P_2 tiene con $\neg vuela(Opus)$ una confirmación parcial, con $\neg vuela(Tweety)$ tiene una anomalía grave, con $vuela(Gringa)$ tiene una observación sorprendente, con $vuela(Pintín)$ tiene una confirmación, y con $vuela(Schmuck)$ tiene otra anomalía grave. Entonces $V_2 = \{\neg vuela(Opus), vuela(Pintín)\}$, $I_2 = \{vuela(Gringa)\}$, $F_2 = \{\neg vuela(Tweety), vuela(Schmuck)\}$.

En estas condiciones, si bien P_1 y P_2 son igualmente progresivos, dado que coinciden sus casos confirmatorios, P_2 es menos exitoso que P_1 porque existe por lo menos una observación ($vuela(Schmuck)$) que es anómala para P_2 pero no para P_1 .

4 La dinámica de los programas

En la Sección anterior identificamos seis posibles resultados de un programa P confrontado a un resultado experimental e_r . En esta Sección trataremos de dilucidar cuál debería ser la estrategia de un programa en función del status que asume frente a nuevos resultados experimentales. Si el programa se confronta con un hecho sorprendente, entonces debería recurrir a algún mecanismo de inferencia ampliativa para poder explicarlo (en este punto es donde se requiere de un contexto de descubrimiento).

En cambio, cuando los resultados son anómalos, entonces es necesario aplicar la heurística negativa, es decir, tratar de modificar las hipótesis periféricas del mismo para

defender el núcleo de la refutación o para mantener progresivo al programa. Cuando existe confirmación total o parcial debería aplicarse la heurística positiva, es decir, tratar de aplicar el núcleo del programa para sistematizar también el cinturón protector, y de esa manera reducir la cantidad de hipótesis *ad hoc*.

Finalmente, cuando el programa está indeterminado por tener una laguna, es decir, si el programa generó teorías o argumentos no definitivos tanto a favor como en contra de la nueva evidencia, entonces es discutible si alguna de las dos heurísticas debería ser aplicada. Una manera de asimilar la importancia de los mecanismos de inferencia que presentamos en los anteriores Capítulos consiste en verlos como posibles formalizaciones de estas estrategias heurísticas de un programa cuando se confronta con una alguno de los casos mencionados.

EJEMPLO 3 *Supongamos estar en la misma situación que en el Ejemplo 1 y deseamos desde el programa P_1 encontrar explicación al hecho sorprendente $vuela(Gringa)$. Una estrategia para explicar un hecho sorprendente e consiste en encontrar por abducción [4, 5] algún hecho particular h el cual, junto con la teoría, permita inferir e .*

En este caso tenemos dos posibles explicaciones más específicas¹ encontradas por abducción: $ave(Gringa)$ y $en-avión(Gringa)$, hipótesis que luego habrá que corroborar por la experiencia. Mientras que estas hipótesis no estén corroboradas, cualquiera de las dos (o ambas) pueden pasar a formar parte del cinturón protector C .

EJEMPLO 4 *Siguiendo nuevamente con el Ejemplo 1 desde el programa P_1 . Tenemos en $\neg vuela(Tweety)$ una anomalía grave. Sin embargo, podemos encontrar una explicación abductiva que corrija la anomalía. En efecto, si agregamos la hipótesis $pingüino(Tweety)$ al programa, entonces el programa tendrá ahora la posibilidad de construir un argumento para $\neg vuela(Tweety)$ que derrota al argumento para $vuela(Tweety)$, es decir, el programa se protege de la observación anómala.*

EJEMPLO 5 *Siguiendo nuevamente con el Ejemplo 1 desde el programa P_1 . Tenemos en $vuela(Schmuck)$ una laguna. En esta situación es necesario o bien establecer un ranking entre los condicionales derrotados utilizados para generar los argumentos para $vuela(Schmuck)$ y para $\neg vuela(Schmuck)$, o bien es necesario modificar alguno de los condicionales (por ejemplo, $pingüino(X) \wedge en-avión(X) \succ vuela(X)$).*

5 Conclusiones y trabajo futuro

Las teorías científicas incorporan un conjunto heterogeneo de conocimiento, jerárquicamente organizado de manera de poder representar una dimensión estratégica. Este aspecto pragmático trasciende al comportamiento puramente lógico, y lo vuelve *racional* en el sentido de incorporar el potencial beneficio del *uso* del conocimiento

¹Es decir, la explicación que realiza la menor cantidad de suposiciones.

por medio de un complejo mecanismo de inferencia denominado vagamente *método científico*.

En este trabajo se estableció un sistema basado en conocimiento inspirado en Lakatos y sus programas de investigación científica. Los programas 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. El sistema permite representar los distintos tipos de conocimiento existentes en una teoría científica, y permite implementar diversos aspectos del razonamiento, especialmente el hecho de que como las algunas leyes son tentativas y sujetas a excepciones y cambios, el razonamiento no puede ser monotónico. También fue considerado un mecanismo de comparación de programas en función del progreso empírico o del grado de confirmación de los mismos frente a nueva evidencia.

Referencias

- [1] Mario Bunge. *La Investigación Científica*. Ariel, Barcelona, 1969.
- [2] Claudio Delrieux. Knowledge Representation and Reasoning in Scientific Research Programmes. In *Proceedings of the Ninth International Conference on Software Engineering and Knowledge Engineering, SEKE 97*, pages 494–501, Madrid, Spain, 1997. Knowledge Systems Institute Press, US. ISBN 0 9641699-5-9.
- [3] Claudio Delrieux. Patterns of Scientific Reasoning as Nonmonotonic Reasoning under Uncertain Evidence. In *Proceedings of the International Conference on Computational Intelligence and Multimedia Applications*, Cairns, Australia, 1997. ICCIMA. ISBN 0 86857 761 8.
- [4] Claudio Delrieux. Computational Theory of Science: Implementing Research Programmes. In *Proceedings of the IC-AI 2000 Conference*, pages 775–783, CSREA Press, ISBN 1-892512-2, 2000.
- [5] Claudio Delrieux. The Rôle of Defeasible Reasoning in the Modelling of Scientific Research Programmes. In *Proceedings of the IC-AI 2001 Conference*, pages 861–868, CSREA Press, ISBN 1-892512-81-5, 2001.
- [6] Claudio Delrieux (editor). Proceedings of the First Workshop on Computer Modelling of Scientific Reasoning and Application. CSREA Press, ISBN 1-892512-73-4. www.lip.uns.edu.ar/cmsra, 2001.
- [7] Claudio Delrieux (editor). Proceedings of the Second Workshop on Computer Modelling of Scientific Reasoning and Application. CSREA Press, ISBN 1-892628-23-7. www.lip.uns.edu.ar/cmsra, 2002.
- [8] Carl G. Hempel. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press, New York, 1965.

- [9] Carl G. Hempel and Paul Oppenheim. The Logic of Explanation. *Philosophy of Science*, 15:135–175, 1948.
- [10] Gregorio Klimovsky. *Las Desventuras del Conocimiento Científico*. A-Z Editora, Buenos Aires, Argentina, 1995.
- [11] Imre Lakatos. *Proofs and Refutations. The Logic of Mathematical Discovery*. Cambridge University Press, 1976.
- [12] Imre Lakatos. *The Methodology of Scientific Research Programmes. Philosophical Papers Vol. I*. Cambridge University Press, 1978.
- [13] Hector Levesque. Knowledge Representation and Reasoning. In J. Traub, B. Grosz, B. Lampson, and N. Nilsson, editors, *Annual Review in Computer Science, Vol. I*, pages 255–288. Annual Reviews Inc., Palo Alto, 1986.
- [14] Karl Popper. *The Logic of Scientific Discovery*. Hutchinson, London, 1959.
- [15] Nicholas Rescher. *Scientific Explanation*. McGraw-Hill, New York, 1969.
- [16] Guillermo R. Simari and Ronald P. Loui. A Mathematical Treatment of Defeasible Reasoning and its Implementation. *Artificial Intelligence*, 53(2-3):125–158, 1992.
- [17] G. A. W. Vreeswijk. Abstract Argumentation Systems. *Artificial Intelligence*, 90(2):225–279, 1997.