Teoría de los Juegos y Selección de Teorías

Fernando Tohmé — Depto. de Economía
Claudio Delrieux — Depto de Ing. Eléctrica
GIIA - Grupo de Investigación en Inteligencia Artificial
ICIC - Instituto de Ciencias e Ingeniería de Computación
Universidad Nacional del Sur
Alem 1253 — (8000) Bahía Blanca — ARGENTINA
claudio@acm.org

Palabras Clave: Inteligencia Artificial; Lógica y Teoría de la Ciencia; Teoría de los Juegos

Abstract

Las teorías científicas pueden considerarse como extensiónes 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*. Sin embargo, es esperable que a partir de un mismo conjunto disponible de conocimiento tentativo, diferentes investigadores adopten distintas jerarquías, por lo que el conjunto de conclusiones a que arriba cada uno es también diferente. Este fenómeno da origen a la coexistencia de dos o más teorías científicas dentro de una misma disciplina, donde cada miembro de la disciplina elije entre las diversas alternativas por medio de algún mecanismo de selección.

En este trabajo se investiga un modelo de elección de teorías en los programas de investigación. En primera instancia se presenta un sistema de representación de conocimiento y razonamiento afín con los *Programas de Investigación Científica* de Lakatos, describiendo la formalización de los distintos procesos de predicción y explicación. Luego presentamos los mecanismos de elección de teorías dentro de un programa de investigación, es decir, el proceso dinámico por el cual las teorías surgen o son abandonadas. El mismo está basado en conceptos de Teoría de los Juegos. Para dichos mecanismos, se estudian las condiciones para la emergencia de equilibrios de Nash.

Teoría de los Juegos y Selección de Teorías

1 Introducción

La implementación de sistemas basados en conocimiento requiere una solución computacionalmente adecuada para los problemas de representación de conocimiento y razonamiento (KR&R). La forma usual y menos problemática de representar conocimiento analítico (universalmente válido), por ejemplo, consiste en utilizar fragmentos decidibles de un lenguaje lógico de primer orden cerrado bajo deducción [11]. Otros *tipos* de conocimiento, por ejemplo lo que no es necesariamente válido, y sus modos asociados de razonamiento, sin embargo, quedan fuera de las posibilidades de esta representación. Un caso paradigmático de la necesidad de incluir estos tipos de conocimiento lo podemos ver en la teoría de la ciencia. La ciencia en desarrollo se desenvuelve por medio de la elaboración más o menos cotidiana de *teorías* por parte de los científicos. Estas teorías científicas, a diferencia de las teorías lógicas, incorporan un conjunto heterogeneo de conocimiento, el cual está jerárquicamente organizado en función del contexto en el que se desenvuelve. Esta organización refleja una dimensión estratégica, contemplando el posible beneficio del *uso* de dicho conocimiento por medio de determinados patrones de inferencia englobados en lo que comunmente se denomina *método*. 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.

En este trabajo desarrollamos como ejemplo ilustrativo una posible formalización de una teoría de la ciencia, inspirada en los *programas de investigación científica* de Lakatos [10]. La formalización presentada en este trabajo comienza por caracterizar distintos *tipos* de conocimiento que se utilizan en una teoría. Este tipo de conocimiento permite representar los resultados experimentales, los postulados hipotéticos y otras piezas de conocimiento tentativo como por ejemplo lo que normalmente constituye el *cinturón protector* de una teoria cientifica [9]. Puede establecerse un ranking de preferencia entre las distintas piezas de conocimiento, reflejando la distinta importancia epistémica que el programa de investigacion asigna a cada una en función de una determinada estrategia. 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. Por lo tanto, existe además un proceso de "selección de

teorías" por medio del cual la comunidad científica pondera las virtudes de cada programa competidor y decide cuál adoptar. La propuesta general de este trabajo es proponer un modelo de selección de teorías basado en nociones de elección social. Esta visión no es nueva, dado que fue primeramente propuesta por Kuhn [8], aunque en términos informales. La evidencia histórica refuerza el punto de vista aquí adoptado. La formalización del mecanismo de selección se establece a partir de "matrices de pago" que reflejan el beneficio de adoptar cada teoría competidora, de una manera abstráctamente similar a la teoría de los juegos. Por último, se demuestran algunas propiedades del modelo, específicamente la existencia de equilibrios de Nash.

2 KR&R y la teoría de la ciencia

Dentro del procedimiento científico, especialmente dentro de las ciencias experimentales, se estableció un conjunto de procedimientos que permiten llevar adelante en forma adecuada la explicación y predicción de fenómenos, la generación de teorías, y la verificación y falsación de las mismas. La formalización de estos procedimientos constituye el objeto de estudio de la teoría de la ciencia. El conocimiento científico se ordena y configura en estructuras complejas. Las unidades de organización 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 (i.e., 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 por lo tanto no tienen el *status* de ser deductivamente válidas. Ésto establece asimetrías en el comportamiento de los patrones de inferencia asociados a estas teorías.

Podemos describir por lo menos tres dominios o niveles de enunciados dentro de una teoría [7]. 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* 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"*. Normalmente los enunciados de este nivel asumen la forma de literales de base, donde tanto los predicados (que representan propiedades, características, etc.) y los términos (que representan objetos, entidades, etc.) son observables. El segundo nivel $\mathcal{N}_2$ está constituido por generalizaciones empíricas o accidentales [12, 5]. 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 (caracterísitica, circunstancia) observable p, normalmente tienen la propiedad q"*. Para representar estos enunciados en un lenguaje lógico, debemos extender al mismo con un "operador condicional" o "implicación *prima facie*" $\succ\!\!-$. Este tipo de

enunciados condicionales, como veremos más adelante, se pueden obtener por abstracción (clases de equivalencia entre enunciados del nivel anterior), o por analogía (isomorfismos con otras estructuras de enunciados). 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 [13].

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 tambien *principios internos*. Este nivel es el más importante de una teoría, pues es el que le confiere su identidad como tal, y permite dar cuenta en profundidad de lo que se conoce en los niveles anteriores. 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.

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. Las primeras descripciones (por ejemplo las del Círculo de Viena) se basaron fundamentalmente en los procedimientos deductivos de la lógica clásica. El esquema subyacente consistía en mostrar que las leyes científicas se infieren de la evidencia. Este acercamiento fue encontrando dificultades insalvables. Hempel [6] fue el primero en proponer que la evidencia debe inferirse de las leyes, y no a la inversa. Según sea que la inferencia se haya realizado antes o después que los hechos deducidos se hayan comprobado, la misma se denomina *predicción* o *explicación*. 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 descubrimiento* 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 ($\mathcal{L} \in \mathcal{N}_2 \cup \mathcal{N}_3$).

La sistematización por medio de este esquema se denominó *paradigma hipotético-deductivo*, dado que el *explanans* constituye una pieza de conocimiento hipotético, del cual se debe deducir la evidencia. El procedimiento de inferir el *explanans* no puede ser deductivo, es decir, $\mathcal{L}$ nunca puede ser *verdadera*. 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. Una conclusión, señalada por Popper [12] 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.

Uno de los criterios pragmáticos expresados en al paradigma hipotético-deductivo consiste en justificar el fracaso de una determinada ley frente a una contrastación dada, no por ser falsa dicha ley, sino por ser inaplicable para ese caso particular. Formalmente

ésto se consigue debilitando el *explanans* con una sentencia particular c, que hace referencia a las condiciones particulares relevantes para la evidencia e a explicar en esta contrastación, donde $c \in \mathcal{N}_1$. Formalmente el esquema deviene en una *implicación contrastadora* $\mathcal{L} \vdash c \Rightarrow e$. En términos de Hempel, el conjunto C de condiciones particulares a los que apela una teoría para efectuar un cubrimiento constituye el conjunto de *hipótesis auxiliares* de la teoría. El cubrimiento por leyes queda formalizado como $\mathcal{L}, C \vdash E$.

Como vimos, mientras una teoría produzca resultados positivos no será completamente abandonada. Este hecho, observado por Lakatos [10], fue el inspirador de su reconstrucción de la dinámica de las teorías científicas, denominadas por él *programas de investigación*. 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 permanecen abiertos y sujetos al cambio y la evolución. Este proceso, sumado a la competencia por la supervivencia de los distintos programas, es totalmente análogo a la evolución natural.

El *núcleo* de un programa es un conjunto de conocimiento que se considera central, y que define la teoría como tal. Este núcleo es el conjunto de conocimientos (leyes, generalizaciones o postulados) que determina la identidad de la teoría y por consiguiente del programa mismo. El núcleo, por lo tanto, se considera definitivo, y el resto de la estructura del programa opera de modo tal de protegerlo de la falsación. Esta protección consiste básicamente en implementar un *cinturón protector* (en la terminología original) de hipótesis auxiliares, que impiden que el núcleo sea falsado. La importancia de todos estos elementos en la dinámica de un programa de investigación puede determinarse solamente dentro del marco selectivo expresado más arriba. Si existen dos o más programas en competencia, probablemente el más exitoso sea aquel cuyo cinturón protector sea menor, aunque las leyes que conforman su núcleo no sean aún totalmente aceptadas. Un ejemplo histórico bien conocido es la crisis de la mecánica newtoniana a principios de siglo. Los experimentos de Michelson-Morley, la mecánica subatómica y ciertos fenómenos astronómicos como la precesión de los equinoccios de Mercurio contradecían las leyes de Newton, y para ser explicados requerían más y más hipótesis *ad-hoc*. La teoría de la Relatividad de Einstein surge como competidora de la mecánica newtoniana, con una ley general que explicaba todos esos fenómenos: la velocidad de la luz es un invariante. Esta teoría fue rápidamente adoptada, pese a que pasaron varios años hasta que se pudo constatar el único fenómeno experimental predicho exclusivamente por la Relatividad: la curvatura del espacio por acción de la gravitación.

3 KR&R y los programas de investigación

El lenguaje de la lógica clásica no es lo suficientemente expresivo como para poder representar formalmente los elementos que constituyen un programa de investigación. Por dicha razón estableceremos en esta sección una serie de puntos de contacto entre el razonamiento científico y el razonamiento no monotónico. Es evidente que el razonamiento científico es no monotónico, porque muy raramente los resultados que produce son firmes. Nuestra descripción formal establece que un programa de investigación está conformado por una *estructura epistémica* compuesta por subconjuntos de los distintos *tipos* de conoci-

miento disponibles, una *estructuración epistémica* que determina la diferente importancia que tiene cada pieza de conocimiento dentro de la teoría, y por un conjunto de *procedimientos de inferencia* que son utilizados en función del contexto en el cual trabaja el programa. Las sentencias que representan conocimiento tentativo general asumen la forma de condicionales o implicaciones *prima facie*. Por ejemplo, la expresión $a(X) >\!\!\!\sim b(X)$ 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 tentativa, se representan como literales indexados l_i, que expresan la disposición a considerar el conocimiento particular l, al provenir éste de un critero tentativo i (información exacta o inexacta, conjetura, criterio estadístico, etc.).

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

- $\mathcal{K}$, *conocimiento lógico-matemático* deductivamente válido;
- $\mathcal{P}$, los *principios internos* de la ciencia en cuestión;
- $\mathcal{H}$, las *hipótesis explicativas* que se derivan de $\mathcal{P}$ y forman parte del *núcleo* de la teoría;
- $\mathcal{G}$, las *generalizaciones accidentales* que surgen como abstracción de un conjunto razonablemente grande de casos particulares, y que también conforman el núcleo de la teoría;
- E, la *evidencia* es el conjunto de datos experimentales que la teoría utiliza;
- C, las *hipótesis auxiliares* utilizadas junto con las hipótesis explicativas para predecir o explicar piezas de evidencia.

DEFINICIÓN **1** *Dado un contexto $\mathcal{K}$, $\mathcal{P}$ (el conocimiento lógico-matemático y los principios internos), una* **Estructura Epistémica** $\mathcal{E}_{\mathcal{K},\mathcal{P}}$ *es una estructura de conocimiento $\mathcal{E}_{\mathcal{K},\mathcal{P}} \subseteq < \mathcal{H}, \mathcal{G}, E, C >$, donde $\mathcal{H}$ es un conjunto consistente de conocimiento intensional tal que $\mathcal{P} \vdash \mathcal{H}$, $\mathcal{G}$ es un conjunto finito de condicionales de la forma $\alpha >\!\!\!\sim \beta$, E es un conjunto conocimiento particular y C es un conjunto de hipótesis auxiliares representadas como conocimiento tentativo de la forma l_i, donde l son literales de base e i corresponde a un criterio de aceptación. Cuando el contexto quede claramente definido, nos referiremos a una estructura epistémica simplemente como $\mathcal{E}$.* □

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

DEFINICIÓN **2** *Dada una estructura epistémica $\mathcal{E}_{\mathcal{K},\mathcal{P}}$ en un contexto $\mathcal{K}$,$\mathcal{P}$, una* **Teoría** $\mathcal{T}$ *es un par $\mathcal{T} = < \mathcal{E}_{\mathcal{K},\mathcal{P}}, \prec >$, donde $\prec$ es un orden parcial sobre los enunciados de $\mathcal{T}$, llamado*

relación de **Preferencia Epistémica**. *$\mathcal{T}$ contiene un elemento $\mathcal{T}_\top$ tal que $\forall\alpha \in \mathcal{T}.\alpha \prec \mathcal{T}_\top$, y un elemento $\mathcal{T}_\bot$ tal que $\forall\beta \in \mathcal{T}.\mathcal{T}_\bot \prec \beta$. De esa manera $\mathcal{T}$ queda reticulada bajo $\prec$.* □

Es importante destacar algunos puntos de estas definiciónes. Cada teoría selecciona enunciados del conjunto total de conocimiento $< \mathcal{H}, \mathcal{G}, E, C >$ que es en principio definible dentro de una disciplina. Estos enunciados, como vimos, no tienen el requisito de ser consistentes entre sí, sino que cada uno de ellos debe ser simplemente consitente con $\mathcal{K} \cup E$. A este subconjunto de enunciados se le asigna *arbitrariamente* un orden parcial $\prec$ que lo estructura. El ordenamiento dependerá estratégicamente del contexto dentro del cual funcione la teoría. De esa manera, en una disciplina pueden coexistir varias teorías, sustentada cada una por una estructura epistémica distinta que la justifica.

El segundo paso en nuestra formalización consiste en definir a un programa como una teoría que progresa en función de determinados procedimientos de inferencia. Uno de los aspectos más importantes consiste en determinar cuál es el conjunto de conclusiones que se justifican a partir de una teoría $\mathcal{T}$, tanto en el contexto de predicción como en el de explicación. Otros procedimientos de inferencia, relacionados con las contrastaciones negativas, la comparación de teorías o la justificación de nuevo conocimiento en función de una "lógica" del descubrimiento, serán presentadas en la sección siguiente. Algunas de las ideas presentadas en esta subsección se basan en el sistema "$\mathcal{P}$" de razonamiento plausible [17, 2], el cual está, a su vez, inspirado en los trabajos de Rescher [14] y Roos [15]. Dado que las teorías no son necesariamente consistentes, la idea esencial es que las conclusiones de una teoría son la consecuencia deductiva de los conjuntos máximamente consistentes de la misma.

DEFINICIÓN **3** *Dada una teoría $\mathcal{T}$ y un subconjunto $T \subseteq \mathcal{T}$, la* **Importancia Epistémica** *de T se define como el conjunto $\{\alpha \in T | \nexists\beta \in T.\beta \prec \alpha\}$ de cotas inferiores de T bajo $\prec$. Dados dos subconjuntos de una teoría T_1 y T_2, diremos que T_1 es epistémicamente más importante que T_2 (denotado como $T_2 \prec T_1$) 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 solución aquí propuesta consiste en considerar la intersección de todos los conjuntos generados bajo distintas extensiones lineales de $\prec$.

DEFINICIÓN **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 **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 **5** *Dada una teoría $\mathcal{T} =< \mathcal{E}_{\mathcal{K},\mathcal{P}}, \prec >$ y una extensión lineal l de $\prec$, un* **Subconjunto Máximamente Consistente** *(SMC) de $\mathcal{T}$ (con respecto al contexto $\mathcal{K}$,$\mathcal{P}$) es un conjunto $\mathcal{E}^l$ que satisface*[1]*:*

[1]Abusando de la notación, utilizaremos subconjuntos de la estructura epistémica como parte del antecedente del operador de consecuencia deductiva clásico, para expresar relaciones de consecuencia que

1. $\mathcal{E}^l \subseteq \mathcal{E}$,
2. $(\mathcal{E}^l \cup \mathcal{K} \cup \mathcal{P}) \nvdash \bot$,
3. $\forall \alpha \in \mathcal{E}^l . \forall \beta \in (\mathcal{E}/\mathcal{E}^l) . \beta \prec \alpha$,
4. $\nexists \mathcal{E}' . \mathcal{E}^l \subset \mathcal{E}' \subseteq \mathcal{E}, (\mathcal{E}' \cup \mathcal{K} \cup \mathcal{P}) \nvdash \bot$.

Es decir, un SMC es (1) un subconjunto de la estructura epistémica $\mathcal{E}$ de la teoría que (2) es consistente con los principios de la misma, (3) incorpora las piezas de conocimiento de mayor importancia epistémica de la teoría, y (4) es maximal en el sentido de que no puede agregársele ninguna pieza de conocimiento de la teoría sin que pierda su consistencia.

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 $< \mathcal{K}, \mathcal{P} >$ y a la relación de importancia epistémica $\prec$). Dicha subteoría está dentro del lenguaje de la lógica clásica. El conjunto $C_{\mathcal{T}}$ de* **Conclusiones** *(predicciones o explicaciones) de una teoría $\mathcal{T}$, entonces, es la clausura deductiva de su subteoría escéptica junto con el contexto, es decir, $C_{\mathcal{T}} = Th(\{\mathcal{K} \cup \mathcal{P} \cup \mathcal{T}_{\not\sim}\})$.* □

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

DEFINICIÓN **6** *Dada una teoría $\mathcal{T} =< \mathcal{E}_{\mathcal{K},\mathcal{P}}, \prec >$ y una consulta q tal que ni $\mathcal{K} \cup \mathcal{P} \vdash q$ ni $\mathcal{K} \cup \mathcal{P} \vdash \neg q$. Entonces definimos:*

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

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

(Aceptación) *q es aceptado si tiene fundamento y no está en duda, o bien si $\mathcal{E}_d \prec \mathcal{E}_f$, es decir, la importancia epistémica de su conjunto de fundamento es mayor que la de su conjunto de duda.*
□

Esta definición es efectivamente computable, ya que incurre en una doble recursión, y las invocaciones a demostrabilidad (Horn), al realizarse por encadenamiento hacia atrás, se computan con técnicas estándar de programación en lógica. El siguiente teorema muestra que el procedimiento de aceptación descripto es correcto y completo con respecto a la definición 5 de conclusiones de una teoría.

TEOREMA **1** *Dada una teoría $\mathcal{T} =< \mathcal{E}_{\mathcal{K},\mathcal{P}}, \prec >$ q es aceptada con fundamento $\emptyset \subset \mathcal{E}_f \subseteq \mathcal{E}$ si y solo si q pertenece a la intersección de todos los subconjuntos máximamente consistentes de $\mathcal{T}$ bajo distintas extensiones lineales de $\prec$.*

---

resultarían si los enuciados de dichos conjuntos subconjuntos fueran utilizados por las reglas de inferencia de dicha relación de consecuencia, particularmente los miembros de $\mathcal{G}$ como implicaciones materiales y los miembros de C como literales.

Demostración
$\Leftarrow$
Si q pertenece al conjunto $C_{\mathcal{T}}$ de conclusiones, entonces $\mathcal{K} \cup \mathcal{P} \cup \mathcal{T}_{\not\perp} \models q$, y por lo tanto, $\mathcal{K} \cup \mathcal{P} \cup \mathcal{E}^l \models q$, 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 $< \mathcal{K}, \mathcal{P} >$) existe algún conjunto de fundamento $\mathcal{E}_s^l$ tal que $\mathcal{K} \cup \mathcal{P} \cup \mathcal{E}_s^l \models q$. 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 $< \mathcal{K}, \mathcal{P} >$ 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 $\mathcal{E}_d^l \prec \mathcal{E}_s^l$ 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 $< \mathcal{K}, \mathcal{P} >$, y, por lo tanto, $(\mathcal{K} \cup \mathcal{P} \cup \mathcal{T}_{\not\perp})$ 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, en toda extensión lineal l se debe cumplir la condición $\mathcal{E}_d^l \prec \mathcal{E}_s^l$. Entonces se sigue que $\mathcal{E}_s$ debe pertenecer a todas las MCS de $\mathcal{E}$ (con respecto al contexto $< \mathcal{K}, \mathcal{P} >$), 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.). □

4 Selección de teorías

La selección de teorías es un proceso de decisión en el cual un conjunto de información es evaluado para elegir un subconjunto que de origen a la teoría preferida. Los resultados de la Sección anterior pueden pensarse como el caso particular en el cual un único agente que toma una decisión en función de la incertidumbre generada por sus fuentes de información. El próximo paso es considerar la existencia de otros agentes que interactúan, los cuales tienen además diferentes preferencias. De acuerdo a nuestro punto de vista, la elección social es el marco formal adecuado para representar el comportamiento colectivo en los aspectos sociológicos de la epistemología y en la historia interna de muchas disciplinas en las que no siempre hubo teorías hegemónicas. Una exposición detallada de los casos históricos más relevantes que apoyan nuestra tesis estaría fuera de lugar en este trabajo. Pueden tomarse como ejemplo la transición de la física medieval a la Newtoniana y luego a la relativista [18], el establecimiento de la teoría de la evolución, y el posterior debate acerca de los equilibrios puntuales [3], y también casos extremos de elección social oficialmente impuesta como por ejemplo la genética de Lysenko.

De acuerdo a la Sección anterior, las preferencias de cada agente constituyen un orden parcial sobre el conjunto de teorías. La elección de la mayoría constituye el "establishment", y si todos adoptan una misma teoría, entonces ésta se convierte en una teoría hegemónica. Cada agente está en condiciones de conocer la preferencia de los demás, y, al mismo tiempo, existe un factor de "presión social" que los condiciona a tratar de adaptar sus propias creencias a las de la mayoría. De esa manera, el proceso de elección social de teorías se puede representar por medio de un juego colectivo donde la matriz de pagos queda determinada por la elección particular de cada agente. El marco formal que representa todas estas intuiciones es el siguiente.

DEFINICIÓN **7** *Dada una estructura epistémica $\mathcal{E}_{\mathcal{K},\mathcal{P}}$ y un conjunto de n agentes $\mathcal{A} = \{\mathsf{A}_i\}, 1 \leq i \leq n$, cada agente A_i tiene una preferencia $\prec_i$ que determina cuál es su conjunto de consecuencias máximamente consistente (SMC) o teoría $\mathcal{T}_i$ elegida. El conjunto $T = \{\mathcal{T} | \mathcal{T}$ es SMC de $\mathcal{E}_{\mathcal{K},\mathcal{P}}\}$ denota todas las teorías (consistentes) posibles, de las cuales, en un determinado momento, cada agente A_i eligió $\mathcal{T}_i \in T$. Denominaremos* **perfil** *$P = (\mathcal{T}_1, \cdots, \mathcal{T}_n) \in T^n$ al conjunto de elecciones simultáneas de los n agentes. De esa manera, existen $|T^n|$ posibles perfiles, cada uno de los cuales brinda apoyo a la elección social de por lo menos una teoría.* □

DEFINICIÓN **8** *Dada una estructura epistémica $\mathcal{E}_{\mathcal{K},\mathcal{P}}$ con su conjunto T de SMC o teorías, y un conjunto de agentes $\mathcal{A}$, cada uno con su preferencia $\prec_i$, entonces el proceso de elección social de teorías puede representarse con un juego $S = \langle \mathcal{A}, T, \{\prec_i\}\rangle$. En S se establece una correspondencia*

$$\rho : T^n \to T$$

tal que una determinada teoría $\mathcal{T}^*$ está en correspondencia con (o condiciona para la elección de) un perfil dado $P = (\mathcal{T}_{i_1}, \ldots, \mathcal{T}_n)$ solo si $\mathcal{T}^*$ es la más frecuente en dicho perfil, es decir

$$\mathcal{T}^* \in \rho(\mathcal{T}_{i_1}, \ldots, \mathcal{T}_n) \text{ solo si para toda } \mathcal{T} \neq \mathcal{T}^*, |\{ttt_i = ttt^*\}| \geq |\{\mathcal{T}_i = \mathcal{T}\}|.$$

□

Hasta aquí no tenemos criterios generales para las preferencias de los agentes por los perfiles $\{\prec_i\}_{i \in N}$, criterios que deben, además, ser distinguidos cuidadosamente de los criterios particulares que utiliza cada agente para elegir su teoría a partir de la información disponible. Para ilustrar este punto con un ejemplo, encontraremos una estrategia que debe seguir un agente para pertenecer a la teoría mayoritaria.

EJEMPLO **2** *Dado un perfil $P = (\mathcal{T}_1, \ldots, \mathcal{T}_n)$ y un agente i, una estrategia que debe seguir i para pertenecer a la teoría mayoritaria, dadas las elecciones conjuntas de los demás agentes $P_{-i} = (\mathcal{T}_1, \ldots, \mathcal{T}_{i-1}, \mathcal{T}_{i+1}, \ldots, \mathcal{T}_n)$, es generar las siguientes preferencias:*

$$(\mathcal{T}_i, \mathcal{T}_{-i}) \preceq_i (\mathcal{T}_i^*, \mathcal{T}_{-i})$$

tal que

$$|\{\mathcal{T}_j = \mathcal{T}_i^*\}_{j \neq i}| > |\{\mathcal{T}_j = \mathcal{T}_i\}_{j \neq i}|$$

o se cumple que

$$|\{\mathcal{T}_j = \mathcal{T}_i^*\}_{j \neq i}| = |\{\mathcal{T}_j = \mathcal{T}_i\}_{j \neq i}| \text{ pero } \mathcal{T}_i \preceq_i \mathcal{T}_i^*).$$

□

Las suposiciones implican que los agentes actúan de acuerdo a las creencias que tienen respecto de las preferencias de los demás. Estas creencias representan las conjeturas que los agentes realizan acerca de la estructura del juego. Dado que normalmente el proceso de discusión científica es público, es justo asumir que las preferencias de cada agente son conocidas por los demás. A partir de todos estos elementos, es posible predecir que la situación determina un equilibrio [4].

DEFINICIÓN **9** *Un perfil* $(\mathcal{T}_1, \ldots, \mathcal{T}_i, \ldots, \mathcal{T}_n)$ *es un* **equilibrio** *de Nash solo si para todo agente* i *no existe* $\mathcal{T}_i^{'}$ *tal que* $(\mathcal{T}_1, \ldots, \mathcal{T}_i, \ldots, \mathcal{T}_n) \prec_i (\mathcal{T}_1, \ldots, \mathcal{T}_i^{'}, \ldots, \mathcal{T}_n)$. □

PROPIEDAD **1** *Dada una estructura epistémica* $\mathcal{E}_{\mathcal{K},\mathcal{P}}$ *con su conjunto* T *de SMC o teorías, y un conjunto de agentes* $\mathcal{A}$*, cada uno con su preferencia* $\prec_i$*, entonces el proceso de elección social de teorías tiene un equilibrio de Nash* $(\mathcal{T}_1, \ldots, \mathcal{T}_n)$ *tal que* $\mathcal{T}_i = \mathcal{T}_j$ *para todo par de agentes* i, j.
DEMOSTRACIÓN*:*
Consideremos un perfil $(\mathcal{T}^*, \ldots, \mathcal{T}^*)$*. Para todo agente* i *se cumple que* $(|\{\mathcal{T}_j = \mathcal{T}^*\}_{j \neq i}| = n - 1$*, es decir, no hay otra teoría* $\mathcal{T}_i$ *tal que* $(|\{\mathcal{T}_j = \mathcal{T}_i\}_{j \neq i}| \geq |\{\mathcal{T}_j = \mathcal{T}^*\}_{j \neq i}|$*. Luego,* $(\mathcal{T}_1, \ldots, \mathcal{T}_n, \ldots, \mathcal{T}_n) \prec_i (\mathcal{T}^*, \ldots, \mathcal{T}^*)$ *para toda teoría* $\mathcal{T}_i$*, lo que por definición significa que* $(\mathcal{T}^*, \ldots, \mathcal{T}^*)$ *es un equilibrio de Nash.*

Ahora supongamos que existe otro equilibrio de Nash $(\mathcal{T}_1, \ldots, \mathcal{T}_i, \ldots, \mathcal{T}_n)$ *tal que para por lo menos dos agentes distintos* i, j *se cumple que prefieren teorías* $\mathcal{T}_i \neq \mathcal{T}_j$ *diferentes. En esta situación pueden darse dos casos:*

- *Existe una teoría* $\mathcal{T}^* \in T$ *tal que* $|\{\mathcal{T}_j = \mathcal{T}^* \neq \mathcal{T}_i\}_{j \neq i}| > |\{\mathcal{T}_j = \mathcal{T}_i\}_{j \neq i}|$.

  Por lo tanto $(\mathcal{T}_1, \ldots, \mathcal{T}_i, \ldots, \mathcal{T}_n) \prec_i (\mathcal{T}_1, \ldots, \mathcal{T}^*, \ldots, \mathcal{T}_n)$*. Pero ésto es absurdo dado que por suposición* $(\mathcal{T}_1, \ldots, \mathcal{T}_i, \ldots, \mathcal{T}_n)$ *era un equilibrio de Nash.*

- *Se cumple que* $|\{\mathcal{T}_j = \mathcal{T}_i\}_{j \neq i}| > |\{\mathcal{T}_j \neq \mathcal{T}_i\}_{j \neq i}|$.

  Entonces, para todo j *tal que* $\mathcal{T}_j \neq \mathcal{T}_i$ *se cumple que* $(\mathcal{T}_1, \ldots, \mathcal{T}_i, \ldots, \mathcal{T}_j, \ldots, \mathcal{T}_n) \prec_j (\mathcal{T}_1, \ldots, \mathcal{T}_i, \ldots, \mathcal{T}_i, \ldots, \mathcal{T}_n)$*. Pero también es absurdo dado que por suposición* $(\mathcal{T}_1, \ldots, \mathcal{T}_i, \ldots, \mathcal{T}_j, \ldots, \mathcal{T}_n)$ *era un equilibrio de Nash.*

Un ejemplo sencillo permitirá ilustrar estos conceptos:

EJEMPLO **3** *Existen tres agentes* i, j *and* k*. Cada uno puede elegir entre tres teorias* $T = \{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3\}$*. Cada agente tiene un orden de preferencias:*

$$i: \quad \mathcal{T}_2 \prec_i \mathcal{T}_3 \prec_i \mathcal{T}_1$$

$$j: \quad \mathcal{T}_3 \prec_j \mathcal{T}_1 \prec_j \mathcal{T}_2$$

$$k: \quad \mathcal{T}_3 \prec_k \mathcal{T}_2 \prec_k \mathcal{T}_1$$

Los órdenes de preferencias entre los perfiles pueden representarse por medio de una función de pago σ*, tal que si para un agente* i *y para los perfiles* P *y* P'*, se cumple que* $P \prec_i P'$*, entonces* $\sigma(P') = 1$ *y* $\sigma(P) = 0$*. De esa manera, se pueden agrupar las preferencias entre perfiles de todos los agentes en una única matriz de pago, de tantas dimensiones*

como teorías elegibles. Cada elemento de la matriz es un vector con el pago que recibe cada agente si un determinado perfil es elegido. El perfil elegido queda determinado por la elección particular de cada agente. La elección de i puede determinar la fila donde se encuentra el perfil elegido, la elección de j determina la columna, y la elección de k la "feta" o capa de la matriz, de atrás hacia adelante. Por ejemplo, la segunda fila de la tercer columna de la primera capa representa el pago por agente del perfil $(\mathcal{T}_2, \mathcal{T}_3, \mathcal{T}_1)$. La matriz de pagos, determinada por las preferencias de cada agente, es:

$$\left[\begin{array}{ccc} (1,1,1) & (1,0,1) & (1,0,1) \\ (0,0,1) & (0,1,0) & (0,0,0) \\ (0,1,1) & (0,0,0) & (0,0,0) \end{array}\right]\left[\begin{array}{ccc} (1,0,0) & (0,1,0) & (0,0,0) \\ (0,0,0) & (1,1,1) & (0,0,1) \\ (0,0,0) & (0,1,1) & (1,0,0) \end{array}\right]\left[\begin{array}{ccc} (1,1,0) & (0,0,0) & (0,0,0) \\ (0,0,0) & (0,1,0) & (0,0,0) \\ (0,0,0) & (1,0,0) & (1,1,1) \end{array}\right].$$

En el ejemplo señalado, el perfil $(\mathcal{T}_2, \mathcal{T}_3, \mathcal{T}_1)$, asigna un pago 0 a todos los agentes. También se observa que los perfiles que son equilibrio de Nash (asigna un pago 1 a todos los agentes) son aquellos en que existe una teoría hegemónica. □

Como muestra este ejemplo, una solución no deseada por la mayoría (en este caso, $\mathcal{T}_3$) puede ser también elegida. Este resultado es indeseado, pero la única manera de evitarlo es por medio de una decisión conjunta, dado que ningún agente tiene incentivo para desviarse individualmente de la decisión de la mayoría. Pero una *coalición*, es decir, un subconjunto de $\mathcal{A}$, puede coordinarse y apartarse conjuntamente del resultado indeseado. La noción de *equilibrio de Nash inmune a las coaliciones* es una propuesta para representar perfiles estables frente a cambios individuales o conjuntos [1].

DEFINICIÓN **10** *Un perfil $(\mathcal{T}_1, \ldots, \mathcal{T}_i, \ldots, \mathcal{T}_n)$ es un equilibrio de Nash* **inmune a las coaliciones** *solo si no existe un subgrupo $\mathcal{A}' \subset \mathcal{A}$ y un perfil $(\mathcal{T}_1, \ldots, \mathcal{T}'_i, \ldots, \mathcal{T}_n)_{i \in \mathcal{A}'}$ tal que $(\mathcal{T}_1, \ldots, \mathcal{T}_i, \ldots, \mathcal{T}_n) \prec_i (\mathcal{T}_1, \ldots, \mathcal{T}'_i, \ldots, \mathcal{T}_n)_{i \in \mathcal{A}'}$.* □

Si proponemos como solución este tipo de perfiles, entonces en el ejemplo anterior (t_3, t_3, t_3) ya no es más una solución estable, dado que la decisión conjunta de i, j y k puede llevarlos a elegir (t_1, t_1, t_1) como mejor solución. Más aún, el otro equilibrio de Nash, (t_2, t_2, t_2) permite una desviación por parte de la coalición formada por i y k, la cual lleva nuevamente a (t_1, t_1, t_1), la cual constituye el único equilibrio de Nash a prueba de coaliciones. Pese a que este modelo de solución parece más adecuado, también tiene la desventaja de que pueden existir casos sin equilibrios inmunes a las coaliciones [2]. Sin embargo, este tipo de situaciones es consistente con los ejemplos históricos, algunos de los cuales fueron mencionados más arriba.

5 Comparación de teorías y la evolución de un Programa

El criterio de comparación de teorías, dentro del comportamiento de un programa, se basa en la relación de importancia epistémica de cada una. Es decir, ante todo las teorías a comparar deben tener una estructura epistémica común (comparten el mismo conocimiento), pero difieren en la importancia que le asignan a cada enunciado. Si las teorías

[2] Por ejemplo, casos análogos al problema de las mayorías cíclicas que son subyacentes a la paradoja de Condorcet [16].

a comparar no tienen la misma estructura, entonces es posible "igualarlas" agregando lo que le falta a cada una en el estrato $\mathcal{T}_\perp$ de menor importancia epistémica. Un ejemplo que muestra cómo se comparan teorías es en el análisis de las distintas alternativas que surgen frente a una contrastación negativa, en una situación que fuerza al programa a evolucionar de manera de asimilar la nueva evidencia.

EJEMPLO 4 *Sea la teoría* $\mathcal{T} =< \{a, a \succ\!\!- b\}, \{\} >$. *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 $\mathcal{T}_1 =< \{a, \neg b, a \succ\!\!- b\}, \{a \prec \neg b, a \prec (a \succ\!\!- b)\} >$. *Según* $\mathcal{T}_1$, *la causa del fracaso es debida a* a, *que no está debidamente justificada, pero* $a \succ\!\!- b$ *se puede seguir manteniendo, es más, crea la presuposición de que* $\neg a$, *la cual habrá que corroborar.*

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

Pueden existir casos en los que se agregan hipótesis auxiliares c *para proteger a la "ley" de ser falsada, creándose una teoría* $\mathcal{T}_3 =< \{a, c, \neg b, a \succ\!\!- b, (a, c) \succ\!\!- \neg b\}, \{\} >$. *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. En esta situación, siempre se desea evitar llegar a la teoría $\mathcal{T}_2$, dado que implica perder una ley constitutiva del programa por causa de un resultado particular, el cual, si bien es evidencia que debe ser tenida en cuenta, puede estar distorsionada por errores metodológicos o sistemáticos, o sujeta a efectos no conocidos.

En los ejemplos históricos, la ley nunca fue abandonada, protegiéndose al programa rechazando los datos y buscando nuevos (caso $\mathcal{T}_1$), o bien, cuando éstos datos se corroboraban en forma abrumadora (algo que llevaría al caso $\mathcal{T}_2$ en forma inevitable), buscando condiciones particulares y nuevas leyes que la completaran (caso $\mathcal{T}_3$). La situación más sencilla, pero la menos progresiva, es intentar ir siempre hacia la situación tipo $\mathcal{T}_1$, y en ese sentido es que estamos en una situación similar a la del Ejemplo 3 donde en primera instancia todos los agentes razonadores eligen esta teoría y el perfil es inmune a coaliciones.

Es importante destacar que aquí existe un aspecto metodológico al que podemos denominar *metaestrategia*, dado que fuerza la elección de una determinada estrategia para defender al programa. En la práctica, ésto puede volverse insostenible, por lo cual es necesario intentar ir hacia la situación $\mathcal{T}_3$, generando nuevo conocimiento, si es que se desea salvar el programa. También es válido remarcar que muchos otros patrones de inferencia (razonamiento hipotético, inducción, abducción o razonamiento por analogía), pueden ponerse en un marco formal como el presentado aquí, aunque por razones de espacio no podemos considerarlos en este trabajo.

6 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 por medio de un complejo mecanismo de inferencia denominado vagamente *método científico*. En este trabajo intentamos mostrar algunos aspectos de la representación de conocimiento y razonamiento dentro de la metodología de la Ciencia, de modo de establecer un sistema basado en conocimiento inspirado en Lakatos y sus programas de investigación científica.

Los mismos fueron caracterizados como una estructura epistémica compuesta por diversos tipos de conocimiento, una relación de preferencia epistémica que los estructura y un conjunto de patrones de inferencia. 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. La relación de preferencia epistémica entre diversas piezas de conocimiento refleja la dimensión estratégica que tienen los programas de investigación.

En la práctica científica usual, especialmente en las ciencias experimentales y aplicadas, es usual la aceptación de patrones de inferencia ampliativos (no válidos) como la formación de conjeturas, inducción, abducción, etc., donde lo que se tiene por *correcto* no proviene de la validez teórica sino de criterios pragmáticos aceptados o históricamente establecidos. De esa manera, se formalizó un programa de investigación como un conjunto de conocimiento estructurado por la relación de preferencia epistémica, el cual progresa por medio de determinados procedimientos de inferencia. En particular, se caracterizaron formalmente los procedimientos de predicción y de explicación en una teoría. A continuación se estudiaron los distintos programas en competencia como un caso de razonamiento multiagente. Se presentaron las condiciones necesarias para establecer una teoría preferida por una comunidad de agentes razonadores, para la cual se demostraron las condiciones de existencia y equilibrio.

Actualmente estamos trabajando en la caracterización de comparadores de teorías que vayan más allá de la relación de preferencia epistémica. Esto significa que deben existir elementos formales (sintácticos) que permitan dar mayor importancia a una teoría por su estructura que por su fundamento, como por ejemplo la especificidad, la presencia de subteorías preferidas, o el uso de mejor evidencia. Por último, queda también abierto el estudio de la *dinámica* de la preferencia epistémica, la cual refleja el cambio de estrategias en el comportamiento progresivo de los programas.

Referencias

[1] B. Bernheim, B. Peleg, and M. Whinston. Coalition-Proof Nash Equilibria: I Concepts. *Journal of Economic Theory*, 42:1–12, 1987.

[2] 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.

[3] Daniel Dennet. *Darwin's Dangerous Idea.* Simon and Schuster, New York, 1995.

[4] D. Fudenberg and J. Tirole. *Game Theory.* MIT Press, Cambridge, Massachussetts, 1991.

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

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

[7] Gregorio Klimovski. *Las Desventuras del Conocimiento Científico.* A-Z Editora, Buenos Aires, Argentina, 1995.

[8] Thomas Kuhn. *The Structure of Scientific Revolutions.* Pitman, London, 1960.

[9] Imre Lakatos. *Criticism and the Growth of Knowledge.* Cambridge University Press, 1970.

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

[11] John McCarthy. Mathematical Logic in Artificial Intelligence. *DÆDALUS, Journal of the American Academy of Arts and Sciences*, 117(1):297–311, 1988.

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

[13] Nicholas Rescher. *Scientific Explanation.* McGraw-Hill, New York, 1969.

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

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

[16] M. Shubik. Game Theory Models and Methods in Political Economics. In K. Arrow and M. Intriligator, editors, *Handbook of Mathematical Economics*, volume 1, pages 19–33. North-Holland, Amsterdam, 1981.

[17] Guillermo R. Simari and Claudio A. Delrieux. Combinanado Plausibilidad y Razonamiento Revisable. In *23 JAIIO, Jornadas Argentinas de Informática e Investigación Operativa*, pages 99–110, Buenos Aires, 1994. .

[18] S. Weinberg. *Dreams of a Final Theory.* Pantheon Books, New York, 1992.