LA CONFIRMACIÓN DE HIPÓTESIS COMO ARGUMENTACIÓN REBATIBLE

Gustavo A. Bodanza
Universidad Nacional del Sur

Resumen

En este artículo se estudia la confirmación de hipótesis como forma de argumentación rebatible del sentido común. Para esto se presenta un sistema argumentativo en el que, a fin de poder a la vez explicar o predecir (razonando de la hipótesis a los *explananda*) y confirmar (razonando de los *explananda* a la hipótesis), se utilizan reglas *default* bidireccionales, *i.e.*, bicondicionales rebatibles. El análisis de tales reglas lleva a plantear la necesidad de contar en el sistema con restricciones contextuales que impidan la confirmación simultánea de hipótesis incompatibles. La justificación de una confirmación, como la de cualquier otra inferencia tentativa del sistema, no se da por medio de reglas lógicas sino a través de la interacción dialéctica de los argumentos rebatibles. Tal interacción permite hallar extensiones del sistema que, según distintos criterios –para los cuales seguiremos a Ph. M. Dung–, determinan la plausibilidad o defendibilidad de las hipótesis.

Abstract

In this article, hypothesis confirmation is studied as a form of common-sense defeasible argumentation. To this aim an argument system is introduced in which bi-directional rules are used –*i.e.*, default biconditionals– in order to enable explanation and prediction (reasoning from the hypothesis to the explananda) but also confirmation (reasoning from the explananda to the hypothesis). The analysis of those rules shows the necessity of adding contextual constraints to the system to prevent the confirmation of incompatible hypothesis. To warrant a confirmation, as any other tentative inference in the system, no logical rules but dialectical notions for the interaction of defeasible arguments are used. This interaction yields extensions of the system that, according to several criteria –which stem from Ph. M. Dung's ideas–, determine the plausibility or defensibility of hypothesis.

Introducción

El propósito de este artículo es procurar un fundamento teórico para el manejo racional de las inferencias confirmativas en un sistema basado en argumentación rebatible. La propuesta recoge la discusión epistemológica sobre el carácter de la confirmación de las hipótesis (Giere [1979], Díez-Moulines [1999], pp. 75-79). Distintos sistemas argumentativos han sido propuestos en el campo de la Inteligencia Artificial en los últimos quince años como estudios de la argumentación rebatible en general (Poole [1985], Loui [1987], Simari-Loui [1992], Prakken-Sartor [1996], Verheij [1996], etc.), pero en ninguno de éstos se ha estudiado la confirmación como argumentación rebatible en particular.

La propuesta es suspender el problema de cómo son las razones que conectan cierta evidencia con la confirmación de una hipótesis, tomando esto simplemente como una licencia inferencial *default* o *prima facie* (*i.e.*, cierta evidencia E implica *prima facie* la verdad de cierta hipótesis H). De este modo, podemos considerar los razonamientos confirmatorios como *argumentos deductivos* en cuanto a su estructura interna (*i.e.*, *prima facie* si E entonces H, y E, por lo tanto H). La primera observación que esto despierta es cómo aceptar tales premisas *prima facie*. Justamente, la propuesta es no preguntarnos por qué la evidencia confirma la hipótesis, sino por qué algunos argumentos confirmatorios a veces son aceptables y otras veces no, dependiendo de qué otros argumentos tengamos en consideración. En la historia de la ciencia suele aparecer como más aceptable la confirmación de una hipótesis que no tiene rivales, como la teoría ptolemaica antes de Copérnico, que la confirmación de otras que compiten por explicar los mismos fenómenos, como en el mismo caso después de Copérnico; también se puede dar la aceptación de ambas hipótesis rivales más o menos igualmente confirmadas, como fue el caso en cierto momento de las hipótesis ondulatoria y corpuscular de la luz, por no haber ventajas teóricas claras y por brindar cada una herramientas de cálculo más sencillas según el fenómeno a explicar o predecir. Si intentamos ver la justificación de los argumentos confirmatorios, no a través de razones inductivas, sino de razones pragmáticas de la argumentación, las preguntas a responder serán: ¿por qué algunos de estos argumentos son más aceptables que otros?, ¿por qué llegan a aceptarse dos hipótesis rivales igualmente confirmadas a un mismo tiempo?, ¿cuál es la racionalidad de tolerar la aparente contradicción?, ¿pueden estas razones dar cuenta de que a veces se preserve una hipótesis refutada por cierta evidencia, o que presenta “anomalías”? Para responder sostendremos que en la pragmática de la argumentación hay ciertas estructuras dialécticas, e intentaremos encontrar un sistema que refleje tales estructuras de un modo general.

El sistema será planteado para el razonamiento de sentido común, y supondremos

que el razonamiento científico podría responder en líneas generales a él si se especificaran ciertas condiciones especiales. La estrategia es la siguiente: mientras lo usual en sistemas de argumentación rebatible para representación del sentido común es utilizar reglas *default* unidireccionales, tales como "si algo es un ave entonces, *prima facie*, vuela", nosotros usaremos reglas bidireccionales tales como "*prima facie*, algo es un ave si y sólo si vuela". El propósito de estas reglas es, además de la aplicación de modus ponens para explicar o predecir (sentido izquierda-derecha), legitimar el razonamiento por la afirmación del consecuente (sentido derecha-izquierda) como licencia argumentativa. Ahora bien, la justificación de argumentos basados en tales *defaults* bidireccionales debe estar regulada por ciertas restricciones que indican incompatibilidades argumentativas con otras creencias contextuales –en las que se incluye una lógica subyacente. Por ejemplo, si contamos con reglas como "*prima facie*, algo es un pingüino si y sólo si no vuela" y "*prima facie*, algo es un ñandú si y sólo si no vuela", para evitar la conclusión "tentativamente, algo es un pingüino si y sólo si es un ñandú", observamos una restricción contextual –que podemos simbolizar con "(pingüino (x), ñandú (x))"– que nos prohíbe predicar "pingüino" y "ñandú" de un mismo individuo x.

Consideraremos las restricciones de este tipo como *estrictas*, en el sentido de que una violación a ellas implica contradicción. También podremos contar con otras restricciones más *débiles*, como por ejemplo (anida-en-árboles (x), anida-en-el-suelo (x), anida-en-las-rocas (x), ...), que servirán para restringir las inferencias tentativas de que un individuo posee más de una de las propiedades en cuestión, pero son débiles en tanto permiten casos excepcionales. Nuestra intención es que el sistema confirme justificadamente una hipótesis cuando se presente como la única o la mejor opción (de acuerdo a un criterio determinado) para explicar ciertos fenómenos conocidos.

La idea de tratar a los *defaults* como reglas bidireccionales surge de un replanteo de la tesis de Giere [1979], que sugiere que la confirmación de una hipótesis H se da por la incorporación de la premisa "si ¬ H, entonces dadas las condiciones antecedentes A muy probablemente no se daría la predicción O"; dado luego O en presencia de A, se infiere que muy probablemente H sea el caso (*ibid.*, p. 96). Nuestra idea es que la modalidad "muy probablemente" es *de dicto*, es decir, opera por fuera del condicional, de modo que éste puede ser reemplazado por su contraposición "si se cumple O, entonces dadas las condiciones antecedentes A muy probablemente sea cierta la hipótesis H". Llevado esto al razonamiento de sentido común, donde las implicaciones contrastadoras que derivamos de una hipótesis no siempre son deductivas, *i.e.*, suelen ser *defaults*, tendremos un *default* para la implicación contrastadora "*prima facie*, si H entonces O" y otro para la implicación confirmativa "*prima facie*, si O entonces H", con lo cual obtenemos un *default* bicondicional "*prima*

facie, H si y sólo si O". Estas reglas, como dijimos, sólo podrán utilizarse bajo determinadas restricciones.

Implementaremos a continuación estas ideas de un modo formal. Definiremos un sistema en el que los argumentos rebatibles se construyen de acuerdo a un conjunto de reglas lógicas que permiten operar con los *default* bidireccionales mencionados, para luego introducir las nociones que determinan el ataque y la derrota entre argumentos. Hecho esto definiremos distintas nociones de "extensión" del conjunto de argumentos del sistema, para lo cual apelaremos a las ideas de Dung [1995]. Las distintas extensiones permitirán representar las diferentes alternativas plausibles de justificación que reciben las hipótesis. Finalmente, mostraremos ciertas proposiciones y ejemplos comparando las extensiones del sistema presentado con otro similar, pero donde operan *defaults* solamente unidireccionales.

Sistema de argumentación confirmativa con restricciones

Representamos la información rebatible de un agente ideal en un sistema argumentativo donde se admite razonamiento hipotético o suposicional. Sea *L* un lenguaje de primer orden. Un *sistema de argumentación confirmativa (con restric - ciones)* es un triplo SAC = K, δ, R ; K Les un conjunto finito y consistente que representa la información no rebatible del agente, que es llamada *el contexto* de SAC, y se divide en dos subconjuntos K = G P, donde las fórmulas de G representan la información general (tal como "los pingüinos son aves"), y las fórmulas de P representan la información particular (tal como "Tweety es un ave"); δ es un conjunto de bicondicionales abiertos llamados *bicondicionales default* o *rebatibles*; R es un conjunto consistente de fórmulas llamadas *restricciones*; cada restricción está representada por una tupla "(P1 (x), ..., Pn (x))", tal que la fórmula representada es verdadera cuando a lo sumo un Pi (1≤ i ≤ n) es verdadero; R está dividido en dos subconjuntos Re y Rd, conteniendo, respectivamente, restricciones *estrictas*, que operan sobre el dominio, y restricciones *débiles*, que operan sobre las inferencias tentativas. Las restricciones sirven para representar incompatibilidades, por ejemplo, "(pingüino (x), ñandú (x))" representa la prohibición estricta de afirmar los predicados "pingüino" y "ñandú" para un mismo individuo "x"; o sea, una restricción estricta (P1 (x), ..., Pn (x)) implica para cualesquiera i, j (1 ≤ i < j ≤ n) y para todo x, {Pi (x) Pj (x)}|– ; en cambio, las restricciones débiles impiden inferencias justificables demasiado crédulas, pero en este caso "P1 (x) ..., Pn (x)" no implica contradicción; por ejemplo, (basquetbolista (x), tenista (x)) es una restricción débil, puesto que "basquetbolista (x) tenista (x)" no implica contradicción, pero si para un x determinado, un argumento rebatible concluye "basquetbolista (x)" bajo el

único supuesto que x está en buen estado físico, y otro concluye "tenista (x)" bajo el mismo único supuesto, esto se interpretará como una incompatibilidad.

Nuestro agente ideal construye argumentos rebatibles con la información en K y δ, mientras R servirá para detectar incompatibilidades argumentativas. La noción de "argumento" que damos a continuación está tomada de Pollock [1990] pero, como se verá, está modificada sustancialmente por la introducción de la regla de "equivalencia rebatible" para la formación de los argumentos.

Definición 1

Un *argumento* es una secuencia de líneas, cada una de las cuales es de la forma *Sup*, *con*, *Def*, {n, m, …} , donde *Sup L* es el conjunto finito de los *supuestos* o *hipótesis* de la línea, *con L* es la *conclusión* de la línea, *Def* d es un conjunto finito de *default*s instanciados en constantes individuales, llamado el *soporte rebatible* de la línea; {n, m, ...} es un conjunto de números naturales correspondientes a los números de las líneas de las cuales la línea actual se deriva directamente. Cuando una línea sea la última de la secuencia, tendremos *Sup*={} en ella[1], puesto que todos los supuestos deberán ser descargados al final del argumento. Los argumentos se construyen de acuerdo a las siguientes reglas:

Creencia básica:

Para cualquier ϕ K, y para cualquier *Sup L*, *Sup*, ϕ , {}, puede ser entrado como una línea de argumento.

Supuesto:

Para cualquier *Sup L* y para cualquier ϕ *Sup*, *Sup*, ϕ, {}, puede ser entrado como una línea de argumento.

Deducción:

Si G {j1, ..., ϕn} |– ϕ y *Sup*1, ϕ1, *Def*1 ... , ..., *Sup*n, ϕn, *Def*n ... ocurren como líneas i1, …, in de un argumento, entonces *Sup*1 ... *Sup*n, ϕ, *Def*1 … *Def*n, {i1, …, in} puede ser entrado como línea posterior.

Equivalencia rebatible:

Para cualquier bicondicional *default* instanciado (ψ ↔ ϕ) δ, si { *Sup*, y [o bien, ϕ], *Def*, ... ocurre como la i-ésima línea de un argumento, entonces *Sup*, ϕ [respectivamente, ψ], *Def* {ψ ↔ ϕ}, {i} puede ser entrado como línea posterior.

Condicionalización:

Si *Sup* {y}, j, *Def*, … ocurre como la i-ésima línea de un argumento, entonces *Sup*, (ψ → ϕ), *Def*, {i} puede ser entrado en una línea posterior.

Otra noción importante es la de *subargumento.* Esta será fundamental para determinar la derrota entre argumentos.

Definición 2

Un argumento σ, cuya última línea es {}, ϕ, *Def*1, ... , es un *subargumento (rebatible)* de un argumento θ, cuya última línea es {}, ψ, *Def*2, ... si y sólo si *Def*1 *Def*2.

Derrota y justificación

Para definir el criterio de derrota necesitaremos de una relación auxiliar, que determine cuándo ocurre un ataque entre dos argumentos. La derrota de un argumento a otro está dada, entonces, por el ataque y la preferencia del primero sobre el segundo. Esta preferencia puede incluir, por ejemplo, la supremacía de los argumentos concluyentes (*i.e.*, construidos prescindiendo de la regla equivalencia rebatible) sobre los argumentos rebatibles, o algún otro criterio como el de especificidad (ver Poole [1985]). Nosotros consideraremos una preferencia abstracta preestablecida que denotaremos con '≥', verificando las propiedades reflexiva y transitiva (*i.e.*, determina un cuasi-orden sobre el conjunto total de argumentos).

Definición 3

Un argumento s, cuya última línea es {}, j, *Def*1, ... *ataca* un argumento θ, cuya última línea es {}, ψ, *Def*2, ... si y sólo si K R {ϕ, ψ} es inconsistente.

Definición 4

Un argumento σ *derrota* a un argumento θ si y sólo si existe algún subargumento θ' de θ tal que (1) σ ataca a θ', y (2) σ ≥ θ'.

Ejemplo 1 (Razonamiento por casos)

Sea SAC:
G = {};
P = { bovino(a) porcino(a) },
δ = { aftosa(x) $\leftrightarrow$ bovino(x),
aftosa(x) $\leftrightarrow$ porcino(x),
aftosa(x) $\leftrightarrow$ ovino(x) }, y
R = Re = { (bovino(x), ovino(x), porcino(x)) }.

(el predicado 'aftosa(x)'significa 'x es transmisor de aftosa'; los demás predicados tienen el significado obvio). El argumento cuya última línea es {}, aftosa(a), {aftosa(a) $\leftrightarrow$ bovino(a), aftosa(a) $\leftrightarrow$ porcino(a)} no tiene atacantes. Supóngase ahora que en lugar de Ptuviéramos P'={aftosa(a)}, entonces el sistema no podría concluir si 'a'es bovino o porcino. Sin embargo, podrán encontrarse los argumentos respectivos en distintas extensiones, como veremos luego, indicando que las dos alternativas son plausibles; en cambio, no habrá extensión que contenga 'ovino(a)', debido a la restricción contextual.

Ahora estamos en condiciones de establecer un criterio de justificación. En este caso seguiremos a Dung [1995], que define una noción bastante general, en el sentido que permite expresar otros enfoques (como las extensiones de las lógicas *default* de Reiter [1980], o los niveles de justificación de Pollock [1990]). El conjunto de sentencias defendibles por el sistema podrá ser determinado según distintos tipos de extensiones, sean éstas de tipo único o múltiple, lo que permite representar distintas intuiciones según los fines del sistema.

Definición 5 (Adaptación de Dung [1995])

Sea S un conjunto de argumentos. Decimos que un argumento σ es *aceptable* en S si y sólo si para todo argumento θ que ataca a σ, hay un argumento υ S tal que υ derrota a θ.

Definición 6

Sea ARG el conjunto de todos los argumentos de un sistema SAC, y sea *f* una función definida por:

1) *f*: $2^{ARG} \rightarrow 2^{ARG}$

2) $f(S) = \{\sigma \mid \sigma$ ARG es aceptable con respecto a S$\}$.

Entonces, un conjunto F ARG de argumentos es una *extensión fundada* de SAC si y sólo si F es el menor punto fijo de *f* –*i.e.*, el menor conjunto F tal que *f*(F) = F. Un conjunto E de argumentos es una *extensión preferida* de un sistema SAC si y sólo si E es un punto fijo de *f* máximamente (c.r. a inclusión conjuntista) libre de conflictos (*i.e.*, no se dan ataques entre sus argumentos)[2].

Las extensiones preferidas pueden ser múltiples, mientras la extensión fundada es única, representando ésta los argumentos más seguros. Esto permite representar distintos usos de las inferencias del sistema.

Lema 1

Sea F la extensión fundada de un sistema SAC, entonces para toda extensión preferida E de SAC se verifica F E.

Prueba. Dado que F y E son puntos fijos, que F es el menor punto fijo de *f*, que *f* es creciente y que determina un orden parcial sobre la clase de todos los puntos fijos, por el teorema de puntos fijos de Tarski –*lattice-theoretical fixpoint theorem*: Tarski [1955, 286-288]– se sigue inmediatamente que F E, para toda extensión preferida E. ■

Definición 7

Una fórmula ϕ L es *plausible* en SAC si y sólo si existe un argumento para ϕ en una extensión preferida de SAC.

Definición 8

Una fórmula ϕ L es *defendible* en SAC si y sólo si existe un argumento para ϕ en la extensión fundada de SAC.

Corolario 1

Toda fórmula defendible en un sistema SAC es plausible en SAC.

Ejemplo 2

En el Ejemplo 1, con P'= {aftosa(a)} la extensión fundada es vacía, con lo cual no

hay fórmulas defendibles, pero tenemos dos extensiones preferidas, una justificando la conclusión plausible 'bovino(a)'y la otra justificando la conclusión plausible 'porcino(a)'.

Ejemplo 3

Este ejemplo es una sobre-simplificación de la rivalidad entre las hipótesis ondulatoria y corpuscular de las radiaciones electromagnéticas, pero permite dar una idea de cómo el sistema podría dar cuenta de una competencia teórica científica. Sean

G =	{};		
P =	{	radiación(a),	
		difracta(a),	
		fotoeléctrico(a)	},
δ =	{	(radiación(x) onda(x)) ↔ difracta(x),	
		(radiación(x) partícula(x)) ↔ fotoeléctrico(x)	}, y
R = Re =	{	(onda(x), partícula(x))	}.

('radiación(x)' significa 'x es una radiación'; 'difracta(x)'significa 'x se difracta'; 'fotoeléctrico(x)'significa 'x manifiesta efecto fotoeléctrico'; los demás predicados tienen el significado obvio.) Entonces tendremos dos argumentos que se atacan, uno que confirma 'onda(a)', y otro que confirma 'partícula(a)'. Esto determina que ésas son dos conclusiones plausibles en distintas extensiones preferidas. La extensión fundada, en cambio, será vacía. Esto muestra que las extensiones preferidas podrían ser aceptables desde un punto de vista metodológico de la ciencia (por ejemplo, para escoger las explicaciones más simples) mientras la extensión fundada podría ser aceptable para encontrar las hipótesis que pueden sostenerse más sólidamente desde un punto de vista teórico. En el ejemplo, que la extensión fundada sea vacía indicaría la falta de una teoría unívoca que dé cuenta del fenómeno particular `a`. Lo mismo ocurriría si en lugar de R = Re tuviéramos R = Rd = {(onda(x), partícula(x))}, como parece más acorde a la teoría cuántica. Sólo que en este caso se admitiría en K la existencia de ciertos elementos `x`, por ejemplo haces de luz, que presentan la dualidad 'onda(x) partícula(x)'. Si el sistema incorpora la regla '(radiación(x) fotón(x)) ↔ (difracta(x) fotoeléctrico(x))', entonces 'fotón(a)' será confirmado y defendible, y además aparecerá justificado junto a la hipótesis plausible 'onda(a)' en una extensión preferida, y justificado junto a la hipótesis plausible 'partícula(a)'en otra extensión preferida.

Comparación con sistemas de *defaults* unidireccionales

Ahora vamos a ver qué resulta al comparar las justificaciones de un sistema basa-

do en *defaults* unidireccionales con nuestro sistema basado en *defaults* bidireccionales. Encontraremos que pueden darse interesantes relaciones entre las extensiones de uno y otro sistema. Sea SAun sistema argumentativo construido como SAC pero cuyos *defaults* son todos condicionales unidireccionales, donde los argumentos rebatibles se forman aplicando modus ponens sobre los *defaults* en lugar de la regla de equivalencia rebatible, y las demás reglas son las mismas que en SAC (el sistema sería muy similar al de Pollock [1990]). Y sea SA$^{\leftrightarrow}$ un sistema SAC cuyo contexto es el resultado de sustituir en SA cada *default* '$\alpha \rightarrow \beta$' por un bicondicional '$\alpha \leftrightarrow \beta$', con lo que cada argumento σ de SA queda traducido en un argumento $\sigma^{\leftrightarrow}$ en SA$^{\leftrightarrow}$, y cada conjunto de argumentos S de SA queda traducido en un conjunto de argumentos S$^{\leftrightarrow}$ en SA$^{\leftrightarrow}$. Es obvio que en SA$^{\leftrightarrow}$ pueden llegar a construirse más argumentos que en SA. Entonces, bajo la condición de que los nuevos argumentos no introducen nuevos pares en $\geq$, y que esa relación se mantiene entre los argumentos traducidos, podemos probar lo siguiente:

Proposición 1

Para cualquier extensión preferida E de SA, existe una extensión preferida E' de SA$^{\leftrightarrow}$ tal que E$^{\leftrightarrow}$ E'.

Prueba. La prueba tiene dos partes, una primera en la que se demuestra que cualquier argumento en E$^{\leftrightarrow}$ pertenece a alguna extensión preferida E' de SA$^{\leftrightarrow}$, y una segunda en la que se demuestra que si dos argumentos pertenecen a E$^{\leftrightarrow}$ entonces ambos pertenecen a una misma E'.

(1) Supongamos que σ E (*i.e.*, $\sigma^{\leftrightarrow}$ E$^{\leftrightarrow}$) pero $\sigma^{\leftrightarrow}$ E'para toda extensión E' de SA$^{\leftrightarrow}$. Entonces se sigue que existe un argumento τ en la extensión fundada de SA$^{\leftrightarrow}$ que derrota a $\sigma^{\leftrightarrow}$. Como τ no es un argumento en SA –de otro modo pertenecería a su extensión fundada también, lo que sería contradictorio con σ E– tenemos que τ viola en SA$^{\leftrightarrow}$ la condición de no alterar el orden establecido por $\geq$, ya que para derrotar a $\sigma^{\leftrightarrow}$ debe ser preferido a éste. Luego, si se cumple la condición de no alterar $\geq$, si σ E entonces σ E' para alguna extensión preferida E' de SA$^{\leftrightarrow}$.

(2) Supongamos ahora que σ y θ son dos argumentos en E tales que $\sigma^{\leftrightarrow}$ y $\theta^{\leftrightarrow}$ no pertenecen a una misma extensión preferida E' en SA$^{\leftrightarrow}$. Entonces $\sigma^{\leftrightarrow}$ y $\theta^{\leftrightarrow}$ se atacan, lo que implica que σ y θ también se atacan, contradiciendo que ambos pertenecen a la misma E.

De (1) y (2) se sigue que para toda extensión preferida E de SA, E$^{\leftrightarrow}$ está contenida en alguna extensión preferida E' de SA$^{\leftrightarrow}$. ■

Corolario 2

Sea F la extensión fundada de SA. Entonces existe una extensión preferida E de SA$^{\leftrightarrow}$ tal que F $\leftrightarrow$ E.

El significado de estos resultados es que en SA$^{\leftrightarrow}$ puede aumentar el número de fórmulas plausibles pero no disminuir, y que todo lo defendible en SAes plausible en SA$^{\leftrightarrow}$. Podemos conjeturar que ciertas condiciones para que lo que serían falacias de afirmación del consecuente en SA, son suficientes para dar conclusiones plausibles en SA$^{\leftrightarrow}$.

Conjetura 2

Si es el caso en SAque
(1) $\alpha \rightarrow \beta$ δ, y
(2) β es defendible, y
(3) $\neg\alpha$ no es defendible,
entonces α es plausible en SA$^{\leftrightarrow}$.

Ejemplo 4

Sea SA:

G = {};
P = {o(a)};
δ = {h(x) $\rightarrow$ o(x), h(x) $\rightarrow$ p(x)};
R = Rd = {(o(x), p(x))}.

SA tiene una sola extensión preferida E cuyo conjunto de consecuencias justificadas es C(E) = Th({o(a), $\neg$ p(a), $\neg$ h(a)}), que coincide con el de su extensión fundada F, C(F) = Th({o(a), $\neg$ p(a), $\neg$ h(a)})[3]. Nótese que respecto de 'h(a)', que estaría tentativamente refutada por '$\neg$p(a)' pero confirmada por 'o(a)', aparece sólo su refutación como defendible. Sea ahora SA$\leftrightarrow$ el resultado de sustituir en SA cada condicional en δ por un bicondicional:

G = {};
P = {o(a)};
$\delta\leftrightarrow$ = {h(x) $\leftrightarrow$ o(x), h(x) $\leftrightarrow$ p(x)};
R = Rd = {(o(x), p(x))}.

SA$\leftrightarrow$ tiene dos extensiones preferidas E1 y E2 cuyos conjuntos de consecuencias

justificadas son C(E1) = Th({o(a), ¬p(a), ¬h(a)}) y C(E2) = Th({o(a), ¬p(a), h(a)}), respectivamente. Entonces se verifica $F\leftrightarrow = E\leftrightarrow$ E1. Por otra parte, '¬h(a)'no es defendible en $SA\leftrightarrow$, pero es plausible al igual que 'h(a)', con lo cual $SA\leftrightarrow$ se muestra más cauto a la hora de refutar. Nótese que en C(E2) se puede tomar 'h(a)' como una explicación plausible en el sistema para 'o(a)', mientras en C(E1) no hay explicación para ese fenómeno.

Conclusión

El sistema SAC muestra, a través de los ejemplos vistos, que la utilización de bicondicionales *default* en un sistema argumentativo permite manejar inferencias de tipo confirmativo con comportamientos aceptables. Por otra parte, la adopción de las ideas de Dung para la definición de extensiones preferidas y fundadas permite considerar las inferencias del sistema según distintos propósitos, como en el ejemplo de las hipótesis ondas/partículas acerca de las radiaciones, donde ciertos argumentos resultarían metodológicamente apropiados, pero no aceptables unívocamente en teoría. En comparación con los sistemas argumentativos basados en *defaults* unidireccionales, el modelo presentado permite extender las justificaciones plausibles para una fórmula a través de las extensiones preferidas.

Pero la importancia mayor de este estudio formal creemos que es filosófica. Aún no sabemos cómo nuestro sentido común se formula hipótesis, pero si suponemos que tal formulación va acompañada de las condiciones bajo las cuales creemos que las hipótesis estarán confirmadas, entonces quizá podamos comprender la racionalidad con que nos valemos de esas creencias. Aquí sostenemos que esa racionalidad es dialéctica, que surge de la interacción de unos argumentos con otros, y que no es reductible a un cálculo como todos los intentos inductivistas suponen.

También dejamos planteados interrogantes acerca de la posibilidad de utilizar el sistema SAC para investigar, en particular, el carácter dialéctico de la competencia teórica científica.

Agradecimientos

Agradezco los comentarios y sugerencias de Fernando Tohmé sobre un borrador de este artículo.

Notas

[1] Para una lectura más clara, al conjunto vacío para sentencias lo denotamos con '{}', mientras al conjunto vacío para números lo denotamos con ' '.

[2] La función *f* y la noción de *extensión fundada* (*grounded extension*) son definidas por Dung [1995]; la noción de *extensión preferida* dada aquí es nuestra.

[3] **Th** (S) es la clausura deductiva de S.

Referencias

Díez, J. A. y C. U. Moulines (1999). *Fundamentos de filosofía de la ciencia.* 2da. ed., Ariel, Barcelona.

Dung, M. (1995). "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games." Artificial Intelligence ,77: 321-357.

Giere, R. (1979). *Understanding scientific reasoning*. Holt, Reinhart and Winston, New York.

Loui, R. (1987). "Defeat among arguments. A system of defeasible inference". Computational Intelligence, 3 (3): 100-106.

Pollock, J. *Nomic probability and the foundations of induction.* Oxford University Press, New York, 1990.

Poole, D. (1985). "On the comparison of theories: preferring the most specific explanation." *Proc. of the Ninth IJCAI,* Los Altos, 144-147.

Prakken, H. and G. Sartor. (1996). "A dialectical model of assessing conflicting arguments in legal reasoning.", Artificial Intelligence and Law 4 (3-4): 331-368.

Reiter, R. "A Logic for Default Reasoning". Artificial Intelligence 13: 81-132, 1980.

Simari, G. and R. Loui (1992). "Amathematical treatment of defeasible reasoning". Artificial Intelligence 53: 125-157.

Tarski, A. (1955). "A Lattice-theoretical Fixpoint Theorem and its Applications". Pacific Journal of Mathematics 5 (2): 285-310.

Verheij, B. (1996). "Two approaches to dialectical argumentation: admisible sets and argumentation stages". Presentado en el *Computational Dialectics Workshop,* junio 3-7, Bonn. Publicado como reporte SKBS/B3.A/96-01.