Significancia de las falacias en los sistemas argumentativos *

Diego C. Martínez † Alejandro García
Grupo de Investigación en Inteligencia Artificial
Departamento de Ciencias de la Computación
Universidad Nacional del Sur
Av. Alem 1253 - (8000) Bahía Blanca - REPUBLICA ARGENTINA
Tel/Fax: (+54)(291)4595135/5136 - E-mail: **{dmartine,ccgarcia}@criba.edu.ar**

Abstract

Entre los problemas que encara la Inteligencia Artificial (IA) se encuentra la representación formal de determinados aspectos del mundo real, como la representación de conocimiento y el razonamiento. En el mundo real, la *argumentación* es utilizada para diferentes propósitos, y en IA han sido propuestos varios modelos formales. Sin embargo, en la argumentación, es posible utilizar argumentos que parecen válidos, pero en los cuales un análisis detallado revela ciertas fallas lógicas. Estos argumentos son denominados *falacias*. Muchos de los sistemas argumentativos propuestos en la comunidad de la Inteligencia Artificial no controlan el uso de argumentos falaces. Algunas de estas falacias están relacionadas con aspectos subjetivos del debate, mientras que otras son intrínsecas de la argumentación, por lo que los sistemas argumentativos deberían aplicar un tratamiento especial a estas últimas. En este trabajo se estudia la posición que adoptan diversos sistemas de argumentación rebatible para resolver este problema.

1 Introducción

Dentro de las Ciencias de la Computación, la Inteligencia Artificial ataca, entre otros, el problema de la representación formal de determinados aspectos presentes en el mundo real. Uno de estos aspectos se refiere al razonamiento y la representación de conocimiento, para lo cual diversos

*Financiado parcialmente por la Secretaría de Ciencia y Técnica, Universidad Nacional del Sur, por la Fundación OSDIC, y la Comisión de Investigaciones Científicas de la Pcia. de Buenos Aires

†Becario de la Comisión de Investigaciones Científicas (CIC)

modelos formales han sido propuestos. Un modelo del razonamiento del sentido común requiere la formalización de procedimientos y principios que caracterizan patrones válidos de inferencia.

La *argumentación* ha sido utilizada en el mundo real para diferentes propósitos en diferentes áreas, y varios modelos de argumentación han sido desarrollados dentro del área de Inteligencia Artificial. Uno de estos modelos, la *argumentación rebatible*, se refiere a un grupo particular de patrones de razonamiento, donde argumentos a favor y en contra de una proposición son producidos y evaluados para verificar la aceptabilidad de dicha proposición.

Cuando se intenta capturar aspectos de un área del mundo real en un sistema formal, es deseable que el formalismo no capture también aquellos problemas inherentes a ése área. Algunos problemas son evitables, o fáciles de descartar, mientras que otros son parte activa del mundo real, por lo cual el modelo debe tratarlos cuidadosamente.

En el caso de la argumentación, es posible que se utilicen argumentos aparentemente válidos, pero que una inspección mas cercana revele alguna falla en su construcción lógica. Estos argumentos se denominan *falaces*. El término *falacia* se utiliza también para referirse a fallas en el proceso que gobierna la exposición de argumentos y contraargumentos durante una argumentación.

Diversos sistemas de argumentación rebatible han sido propuestos. Dung [DUNG93] propuso un sistema argumentativo general, donde se abstrae de los conceptos de argumento y derrota. Prakken y Sartor [PRAK] desarrollaron un sistema que utiliza el lenguaje de la programación en lógica extendida, al igual que García y Simari [AG97], éste último haciendo una clara distinción entre reglas utilizadas para representar el conocimiento estricto y tentativo. Otros desarrollos sobre argumentación rebatible pueden encontrarse en Vreeswijk [VREE97], Bondarenko,Toni y Kowalski [BOND93, BOND97], Francesca Toni [KOWA96], Bart Verheij [BART] y Ronald Loui [LOUI97]. El lector interesado puede dirigirse a los siguientes *surveys* en argumentación rebatible: Prakken y Vreeswijk [PVREE], y Chesñevar, Maguitman y Loui [CML98].

El estudio de las falacias en la argumentación rebatible permitirá evitar procesos de razonamiento incorrectos, que no incurran en las mismas falacias del mundo real. De esta forma, aquellos sistemas de argumentación rebatible que incorporen un tratamiento de las falacias, permitirán modelar un proceso de argumentación sin errores.

El objetivo de este trabajo es estudiar el tratamiento que los sistemas de argumentación rebatible deben dar a las falacias. Para ello se presentarán diversas falacias que han sido reconocidas y estudiadas fuera del ámbito de la Inteligencia Artificial, y luego se mostrará el tratamiento que es dado por algunos de los sistemas de argumetación rebatible existentes a estas falacias. También se mostrará aquellas falacias que no son tratadas, indicándose los problemas que ello acarrea. En la sección siguiente se enumeran las características comunes a todos los sistemas

de argumentación rebatible. En la sección 3 se presentan las falacias, y se establecen aquellas que son relevantes para estos sistemas. Finalmente, se estudian tres casos reales, y su posición frente a las argumentaciones falaces.

2 Elementos de un sistema de argumentación rebatible

Como fué expresado por Prakken en [PVREE], los sistemas de argumentación rebatible cuentan en general con los siguientes cinco elementos, aunque algunas veces implícitos:

- Un lenguaje lógico subyacente, utilizado para representar el conocimiento.
- Una noción de argumento.
- Definiciones de conflictos entre argumentos.
- Un mecanismo de inferencia que permita establecer si un argumento es aceptado.

Los sistemas argumentativos son entonces construídos en base a un lenguaje lógico, y una noción de consecuencia lógica, definiendo así el concepto de argumento.

Por ejemplo, podría utilizarse un lenguaje de primer orden, o un lenguaje más restrictivo, como las cláusulas de Horn de la programación en lógica, utilizando el operador de consecuencia respectivo a fin de obtener una derivación que permita la construcción de argumentos. Aunque estos conceptos se encuadran con las características convencionales de un sistema lógico, los restantes tres elementos son inherentes a la argumentación rebatible.

La noción de *conflicto* entre argumentos debe establecer cuándo un argumento ataca o contradice a otro. Por ejemplo, dos argumentos estarán en conflicto si sustentan respectivamente conclusiones contradictorias. El concepto de conflicto entre argumentos no involucra la comparación entre argumentos. Sin embargo, si se provee tal criterio de comparación, puede entonces definirse una relación binaria de *derrota*. De esta forma, un argumento derrotará a otro si están en conflicto y el primero es mejor que el segundo de acuerdo al criterio de comparación utilizado. Es importante destacar, que la relación de derrota sólo establece una relación de fuerza entre dos argumentos en conflicto. Sin embargo, en el proceso de argumentación, pueden estar involucrados más de dos argumentos y la relación de derrota no alcanzará para decidir la aceptabilidad de un argumento.

La aceptabilidad de los argumentos depende de la interacción de todos los argumentos disponibles al momento de la argumentación. Es por este motivo que el proceso de argumentación es formalizado en algunos sistemas mediante un análisis dialéctico, mientras que en otros como una disputa entre el proponente de una conclusión y su oponente.

3 Falacias en la argumentación

Un argumento puede ser falaz por diferentes motivos. En esta sección se presentarán varios tipos de falacias. Las primeras se refieren a aspectos subjetivos. Las segundas aparecen como consecuencia de una mala construcción de los argumentos.

3.1 Falacias de Relevancia

Existen falacias [ATWEB, YOLDER99, EREAS] que se refieren a aspectos subjetivos del debate o de la forma en que los argumentos son enunciados. Algunas de ellas están fuertemente relacionados con el contexto de una discusión real, y no son mayoritariamente de nuestro interés. Los siguientes tipos de argumentos, denominados *falacias de relevancia*, intentan sustentar una conclusión con algo más que la evidencia real.

- *Argumentum ad hominem*: argumento donde se ataca el carácter o la personalidad del oponente, en lugar de su tesis. (*Juan es el ladrón, porque ya ha robado con anterioridad*).
- *Argumentum ad misericordiam*: argumento en el que se intenta persuadir apelando a la piedad del oponente. (*No me culpen de robo, sufro mucho porque en mi casa no hay dinero.*)
- *Argumentum ad verecundiam*: argumento en el que se intenta persuadir apelando a una autoridad. (*Dios existe, porque Isaac Newton creía en Él*).
- *Argumentum ad baculum*: argumento en el que se intenta persuadir a través del miedo o la amenaza de fuerza.(*Vive bajo las reglas de Dios, o te quemarás en el infierno.*)
- *Non causa pro causa*: donde se afirma que algo es la causa de un evento, aún cuando no está fehacientemente demostrado.
- *Plurium interrogationum*: donde se pretende una simple respuesta a una pregunta en realidad compleja.
- *Reification*: donde se comete el error de tratar un concepto abstracto como algo concreto.

Todos ellos aparentan ser argumentos válidos, pero una inspección más cercana revela sus fallas lógicas. En algunos de ellos la falacia radica en que suelen presentar premisas que realmente no son correctas, en el sentido de que no tienen relación directa con la conclusión, o bien apelan a aspectos que escapan del tema principal del debate. Por ejemplo, el siguiente argumento

> "Yo merezco el trabajo, *porque* estoy calificado: tengo experiencia y realmente necesito el dinero"

es un argumento *ad misericordiam.* Las razones expuestas en él no son razones válidas para la conclusión: entre ellas existe una apelación a la piedad del empleador.

El problema con esta clase de argumentos es que las conclusiones basadas en ellos no están basadas en la verdad, sentido común o lógica sino en emociones que favorecen a aquellos que enuncian los mejores argumentos.

En el intento de construir un modelo formal de un aspecto de la realidad, es importante tener en cuenta no trasladar a dicho modelo aquellos problemas (como las falacias) presentes en el mundo real.

Los sistemas de argumentación rebatible no deberían presentar argumentaciones falaces. Como las falacias mencionadas anteriormente tratan con aspectos emocionales, éstas no aparecen en la mayoría de las formalizaciones de la argumentación rebatible, pues no pueden ser capturadas por el modelo formal.

En el caso en que efectivamente el sistema contemplara correctamente estos aspectos en la representación de conocimiento, bastaría que el sistema reconozca y evite la falacia.

3.2 Falacias en la construcción de un argumento

Por otro lado, algunas falacias son verdaderamente interesantes, y no tienen relación con aspectos subjetivos de la discusión, sino con la estructura y el flujo de la misma. Por ejemplo:

- *Argumentum ad Ignorantiam*: es cualquier argumento que usa la regla: " Si algo no puede ser probado incorrecto, entonces debe ser cierto" o "Si algo no puede ser probado cierto, entonces debe ser incorrecto".

Este modo de razonamiento utiliza la ausencia de evidencia de la verdad (o falsedad) de una proposición como prueba de lo opuesto. Por ejemplo, *si no hay evidencia de vida extraterrestre, entonces no existe.* Esto es tratado como falacia por muchos filósofos, pero este modo de razonamiento es el núcleo de muchos sistemas de razonamiento no monotónico. Desde este punto de vista, la *negación por falla* en la programación en lógica sería una falacia, de la misma forma que la justificación de argumentos, pues los argumentos se vuelven justificaciones ante la ausencia de información contradictoria (mejores contraargumentos).

Sin embargo, no hay nada incorrecto en estos sistemas, ya que depende del nivel de detalle con el que se trata en el mundo real, y de las herramientas utilizadas en éstos modelos. En los sistemas no monotónicos, es correcto aceptar una conclusión cualquiera, basada en un argumento *ad ignorantiam*, debido a que su base de conocimiento nunca es completa.

Las falacias *petitio principii* ("que evaden los hechos"), presentan mayor interés para el análisis de los sistemas de argumentación rebatible. Pueden presentarse de diferentes formas:

1. *Ofreciendo, como premisa una simple redefinición de la conclusión deseada*
2. *Un argumento circular, más complejo que (1) pero que eventualmente pretende justificar la conclusión con la conclusión misma.*

Un argumento de este tipo se considera *circular* ya que la conclusión a ser probada se utiliza como una premisa para construir el argumento. Por ejemplo: *"Dios existe, pues Dios hizo que la Biblia fuera escrita, y la Biblia afirma que Dios existe"*. Algunos autores también reconocen esta falacia bajo el nombre de *circulus in demonstrando*. La argumentación circular es soprendentemente común.

Otras falacias interesantes son los *Argumentum non sequitur*. En esta clase de argumentos la conclusión simplemente falla como consecuencia de las premisas. Pueden también presentarse de diferentes formas:

- *Affirming the consecuent*: cualquier argumento de la forma: Si A entonces B; B; por lo tanto A.
- *Denying the antecedent*: cualquier argumento de la forma: Si A entonces B, No A, por lo tanto No B.
- *Inconsistency*: sustentar que afirmaciones contrarias o contradictorias son ambas verdaderas.
- *Stolen Concept*: usar un concepto para atacar un concepto del cual lógicamente depende.

Este tipo de falacias es tan común que utilizan el término *falacia* para referirse sólo a éstas cuatro formas.

El primero es simplemente un uso incorrecto de la lógica. *Negar el antecedente* suele ser usado como justificación de lo contrario de una proposición determinada, cuando en realidad el efecto de dicha acción es simplemente la no justificación de esa conclusión. Por ejemplo, el argumento

Si María me ama, entonces María se casará conmigo

es derrotado por el argumento

María no me ama

por lo que la conclusión "María se casara conmigo" no está justificada. Esto no significa que efectivamente María *no* se casará conmigo. Puede haber otras razones por las cuales María se case conmigo.

En los sistemas argumentativos en los cuales la derrota de los argumentos es activada por inconsistencia, los argumentos que se derrotan a si mismos son argumentos *non-sequitur*, pues son argumentos que presentan inconsistencias en su estructura. También, los sistemas de argumentación rebatible suelen incurrir en una *argumentación contradictoria*, donde un argumento es usado para defender y atacar la misma conclusión, lo cual puede verse como una forma de falacia *stolen concept*.

4 Sistemas argumentativos y falacias

En esta sección, se estudian diferentes sistemas argumentativos existentes, en función de las falacias argumentativas arriba mencionadas. En el siguiente apartado expandimos el ámbito de las falacias no solo a argumentos individuales, sino también al análisis dialéctico requerido para justificar un argumento.

La mayoría de las situaciones falaces definidas en la seccion anterior, están basadas en la estructura de los argumentos. Sin embargo, estos conceptos pueden ser aplicados a cualquier debate completo, no sólo a argumentos individuales. Cualquier proceso de justificación involucra la comparación de argumentos y contraargumentos. Dicho proceso puede ser visto como una discusión entre un proponente de una conclusión C y su oponente. Los argumentos del proponente son la posible explicación para la conclusión C, mientras que los argumentos de oponente son la posible explicación para la negación de C. De esta forma, algunas falacias pueden ser encontradas en el proceso dialéctico, tales como *petitio principii* y *Stolen concept*.

4.1 Dung

Argumentation Frameworks

Phan Minh Dung define un sistema argumentativo abstracto [DUNG93], donde deja la estructura de los argumentos completamente sin especificar, centrándose en los conflictos existentes entre

ellos. Asume la existencia de un conjunto de argumentos ordenados por una relación binaria de derrota. Define un *argumentation framework* como un par $\langle AR, attacks \rangle$ donde AR es un conjunto de argumentos y *attacks* es una relacion binaria sobre AR, que indica los conflictos entre los argumentos. Por ejemplo,

Ejemplo 4.1. *AF=$\langle AR, attacks \rangle$ es un argumentation framework tal que*

$$AR = \{A, B, C\}$$
$$attacks = \{(A, B), (B, C), (C, A)\}$$

Dada la simplicidad del argumentation framework, se puede usar un grafo dirigido para utilizar una forma gráfica para denotar el estado de los conflictos.

El concepto principal es el de *aceptabilidad de argumentos*. Un argumento A es aceptable con respecto a un conjunto S de argumentos si todo argumento que derrota a A es derrotado por algún argumento de S. En base a ésto, define la semántica del *argumentation framework* a través del punto fijo de una función monótona. Dado que en este sistema no está definido un lenguaje para la representación de conocimiento no existe en él una noción concreta de argumento y derrota. Debido a este alto nivel de abstracción, las únicas falacias que pueden presentarse son aquellas que surgen de la interacción de los argumentos, como la argumentación circular y contradictoria. Las falacias como los argumentos *Non sequitur* sólo tienen sentido cuando hablamos de la estructura de los argumentos, aunque Dung admita la posibilidad de argumentos que se derrotan a sí mismos.

El *argumentation framework* del ejemplo 4.1 presenta un caso de argumentación contradictoria [DCM99], donde los argumentos atacan y defienden indirectamente a los restantes. Este argumentation framework no es *bien fundado* y es *controversial* [DUNG93]. El hecho de que, por ejemplo, el argumento B sea atacado *indirectamente* por B, implica que B es un argumento *stolen concept*, pues defiende un concepto mientras durante la discusión lo ataca. La semántica escéptica de los argumentation frameworks define a los argumentos de este tipo como argumentos *no justificados*, lo cual es etimológicamente correcto. La argumentación circular también está presente en este sistema, lo que provoca una falacia argumentativa "*petitio principii*".

Ejemplo 4.2. *Sea AF=$\langle AR, attacks \rangle$ un argumentation framework tal que*

$$AR = \{A, B, C, D\}$$
$$attacks = \{(A, B), (B, C), (C, D), (D, A)\}$$

En este caso, un argumento, por ejemplo A, es utilizado en la discusión dialéctica para defender a A. Esto puede verse como que entre las explicaciones para justificar A, enunciamos

el argumento A. Nuevamente, la semántica de los argumentation framework clasifica a estos argumentos falaces como *no justificados*.

Una característica importante en este sistema es que la misma semántica esceptica definida por Dung incurre en una clase particular de falacias denominada "*Argumentum ad logicam*", donde se establece que una proposición es falsa porque ha sido presentada como la conclusión de un argumento falaz. En [DCM99] se muestra que en el siguiente ejemplo

Ejemplo 4.3. *el argumentation framework* $AF{=}\langle AR, attacks\rangle$ *donde*

$$AR = \{A, B, C, D\}$$
$$attacks = \{(A, B), (B, C), (C, D), (D, B)\}$$

el argumento A, que no participa en el ciclo de argumentación contradictoria, es clasificado también como argumento *no justificado* [1].

Programas Lógicos extendidos con negación explícita

Posteriormente Dung definió una extensión a la programación en lógica convencional en la cual se permite la utilización de negación explícita [DUNGLP]. Establece que por cada programa extendido se puede construir un argumentation framework cuya semántica define la semántica del programa.

Los literales de los programas extendidos pueden ser literales *objetivos* o literales *subjetivos*. Un literal *objetivo* es un atomo A o la negacion explicita $\sim A$ de un atomo A. Un literal *subjetivo* es un literal de la forma "not L" donde L es un literal objetivo. Una cláusula de programa extendido es de la forma $L_0 \leftarrow L_1, ..., L_n, notL_{m+1}, ..., notL_{m+n}$" donde los L_i son literales objetivos. Los literales subjetivos que aparecen en la prueba son las hipótesis sobre las cuales la prueba es construída. La aceptación de la prueba depende, entonces, de la aceptación de las hipótesis. Dung define, para un programa P, un *argumento* con respecto a P, como un conjunto de literales subjetivos. Un literal objetivo L es *soportado* por un argumento A si existe una prueba para L cuyas hipótesis están contenidas en A.

En estos progamas pueden existir argumentos que se derroten a sí mismos. En un programa P, un argumento A se derrota a sí mismo si existe un literal L tal que A soporta a L y $\sim L$. En este caso, el argumento A es un argumento "*non sequitur*" al ser inconsistente, puesto que soporta dos conclusiones contradictorias. Sin embargo, estos argumentos son excluídos del argumentation framework para el programa P, y ningún argumento *non sequitur* es un argumento justificado.

[1] Hay que tener en cuenta que el motivo de la discusión aquí es la justificación o no del **argumento**

4.2 Prakken y Sartor

Motivados por el razonamiento legal, Henry Prakken y Giovanni Sartor [PRAK] definen una extensión de la programación en lógica, con prioridades rebatibles, basada en argumentos.

En este sistema, un literal fuerte es un átomo A, o un átomo negado $\neg A$. Se denota a $\overline{A}$ al complemento del literal A. Un literal débil es un literal de la forma $\sim L$ donde L es un literal fuerte. Informalmente, $\sim L$ significa "no existe evidencia para L". Una *regla* es una expresión de la forma

$$\text{r: } L_0 \wedge ... \wedge L_1 \wedge \sim L_k \wedge ... \wedge \sim L_m \implies L_n$$

donde la conjunción a la izquierda de la flecha es el antecedente y L_n es el consecuente. Existen dos tipos de reglas: las reglas *rebatibles* y las reglas *estrictas*. Solo las reglas rebatibles pueden contener literales débiles. El conjunto de estas reglas, además, esta ordenado por un orden parcial.

En este sistema un argumento es un conjunto ordenado de reglas tal que todo literal fuerte en el antecedente de una regla, es también el consecuente de alguna regla anterior. Un literal L es una *conclusión* de un argumento A si L es el consecuente de alguna regla de A. Un literal L es una hipótesis de un argumento A, si $\sim\overline{L}$ ocurre en alguna regla de A. Existen dos formas de conflictos entre argumentos: *rebut*, donde dos argumentos permiten alcanzar conclusiones contradictorias, y *undercut* donde un argumento tiene como conclusión el complemento de la hipótesis de otro. Sea $CONC(A)$ el conjunto de conclusiones de un argumento A; $ASS(A)$ el conjunto de hipótesis de un argumento A, y sea S el conjunto de reglas estrictas. Un argumento A ataca a B, si

- $CONC(A) \cup CONC(B) \cup S \vdash \perp$ (Rebuts)
- $CONC(A) \cup S \vdash \neg\varphi$ para cualquier $\varphi \in ASS(B)$ (Undercuts)

Finalmente, un argumento A derrota a un argumento B si

- A "undercuts" B , o
- A "rebut" B, pero B no "undercut" A

La semántica de este sistema está basada en la semántica para argumentation frameworks de Dung, y por lo tanto padece los mismos problemas, en cuanto a argumentación "stolen concept" y argumentación "petitio principii". En este sistema pueden encontrarse también argumentos falaces del tipo "non sequitur": los argumentos pueden tener conclusiones contradictorias, por

lo que un argumento que posee como conclusiones a A y $\neg A$ se derrota a si mismo. Nuevamente, el proceso de justificación también recurre en una falacia "*Argumentum ad logicam*" al clasificar como no justificados a argumentos que no pertenecen a algun ciclo argumentativo, pero que sin embargo son derrotados por un argumento "non sequitur" o un argumento "petitio principii", al igual que en Dung.

4.3 Programas Lógicos Rebatibles

Definida como una extensión de la programación en lógica convencional, la *Programación en Lógica Rebatible* [SIM92, AG97] permite capturar aspectos del razonamiento del sentido común, difíciles de capturar en otras extensiones. Los programas lógicos rebatibles (PLR) permiten representar información rebatible y no rebatible, asi como también información incompleta y potencialmente inconsistente. Están formados por dos tipos de reglas, las reglas *fuertes* (de la forma "Cab $\leftarrow$ Cuerpo"), utilizadas para representar conocimiento estricto, y las reglas *rebatibles* (de la forma "Cab $\prec$ Cuerpo") utilizadas para la representación de conocimiento tentativo. El conjunto de reglas fuertes, denotado S, debe ser consistente, mientras que el conjunto de reglas rebatibles, denotado D, puede ser inconsistente.

Debido a que es posible derivar literales complementarios, los PLR utilizan a la argumentación rebatible para poder decidir entre metas contradictorias. Un argumento para un literal q se obtiene de una derivación de q, que sea consistente y minimal. Los argumentos se denotan con un par $\langle$ A,h $\rangle$ donde A es el conjunto de reglas rebatibles utilizadas para derivar la *conclusión* h. Un argumento $\langle$ A,a $\rangle$ derrota a un argumento $\langle$ B,b $\rangle$ cuando existe un subargumento $\langle$ B_1,b_1 $\rangle$ de $\langle$ B,b $\rangle$ tal que $\langle$ A,a $\rangle$ esté en *desacuerdo* con $\langle$ B_1,b_1 $\rangle$, esto es $S \cup \{a, b_1\} \vdash \perp$. Una consulta q tendrá éxito cuando el argumento A que la sustente sea una justificación.

Para establecer que A es un argumento derrotado, se deben construir contraargumentos, que serán sus posibles derrotadores. Como los derrotadores son a su vez argumentos, pueden existir derrotadores para los derrotadores, y así sucesivamente, lo que requiere un análisis dialéctico completo. Una línea de argumentación, es una sucesión de argumentos $[A_0, A_1, ..., A_n]$ donde figuran alternadamente argumentos a favor y en contra del primero de la secuencia (A_0). Los argumentos A_{2k+1} son denominados *argumentos de interferencia* para A_0, y los argumentos A_{2k} son denominados *argumentos de soporte* para A_0.

En los programas lógicos rebatibles los argumentos son derivaciones consistentes, es decir, dado un argumento A, se cumple que $S \cup A \nvdash \perp$. Por lo tanto en este sistema no se encuentran argumentaciones falaces "*non sequitur*" del tipo *inconsistente*. En los PLR no existen los argumentos que se derrotan a sí mismos.

Lo interesante de este sistema es que en el proceso dialéctico de justificación de argumentos pueden encontrarse, al igual que los sistemas anteriores, falacias "*petitio principii*" y "*stolen concept*", pero sin embargo existe un tratamiento que permite clasificar a los argumentos falaces como *no justificados*.

En los programas lógicos rebatibles se requiere que todas las líneas de argumentación existentes para cualquier argumento cumplan las siguientes condiciones:

1. Los conjuntos de argumentos de soporte y de interferencia deben ser, cada uno, un conjunto concordante de argumentos. Un conjunto de argumentos $\{A_1, A_2, ..., A_n\}$ es concordante si $\bigcap_i A_i \cup S \nvdash \bot$
2. Ningún A_i es subargumento de algún A_k $(k \leq i)$

De esta forma, los argumentos involucrados en una falacia *petitio principii* o *stolen concept* poseen una línea de argumentación finita a pesar del ciclo, y son clasificados como argumentos *no justificados*.

Mas aún, este tratamiento evita caer en una falacia "*Argumentum ad logicam*". Un argumento que no forma parte de una falacia *stolen concept* o *petitio principi* pero que sin embargo es derrotado por un argumento falaz, no necesariamente es un argumento no justificado, lo que sí ocurre en los sistemas de *Dung* y *Prakken y Sartor*. En los PLR, ante una situación como la del ejemplo 4.3, el argumento A está justificado, pues esta derrotado por un argumento *falaz*, y ésa es la razón de su justificación, lo cual es lógicamente correcto.

5 Conclusiones

En este trabajo se revisaron algunas de las falacias argumentativas más importantes. Algunas de ellas, de carácter subjetivo como *argumentum ad hominem* o *argumentum ad misericordiam* son inevitables en los modelos formales y pueden encontrarse en prácticamente todos los sistemas argumentativos. No constituyen un defecto en sí, pues dependen de la temática y no necesariamente del proceso de discusión.

Otras falacias, como *non sequitur* son frecuentes en algunos sistemas de argumentación rebatible que permiten la construcción de argumentos capaces de derrotarse a sí mismos. Esta es una característica no deseable. Algunos sistemas [AG97] evitan incluir esta clase de argumentos y otros tratan con ellos de manera tal que, apelando al sentido común, no sean argumentos justificados [DUNG93, DUNGLP, PRAK].

Las falacias existentes en prácticamente todos los sistemas de argumentación rebatible, son las del tipo *petitio principi*, donde se produce un ciclo en la argumentación, y *stolen concept*

donde se incurre en una contradicción en el proceso dialéctico. Estas falacias son intrínsecas de la argumentación y por lo tanto, heredadas por los sistemas argumentativos. Los argumentos falaces de este tipo no deben ser justificaciones. Los sistemas de Dung [DUNG93, DUNGLP], Prakken y Sartor [PRAK] y García y Simari [AG97] controlan este tipo de falacias. Finalmente, y como efecto colateral de estas falacias, muchos sistemas incurren en una *Argumentum ad logicam*, al clasificar como argumento no justificado a aquellos que son atacados por alguna falacia *petitio principii* o *stolen concept*. Este problema no es tenido en cuenta en muchos sistemas. En la Programación en Lógica Rebatible [AG97, SIM92] esto no sucede, precisamente porque existe, a diferencia de los demás, un tratamiento específico para esta clase de falacias [GCS94, DCM99].

Bibliografía

[AG97] Mg. Alejandro García. *La programación en lógica rebatible. Su definición teórica y computacional* Tesis de Magister en Cs. de la Computación. Universidad Nacional del Sur. 1997.

[ATWEB] *The Atheism Web: Logic and Fallacies.* http://www.infidels.org/news/atheism/logic.html

[BART] Bart Verheij *Rules, reasons and arguments: formal studies of argumentation and defeat.* ISBN 90-9010071-7

[BOND97] A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

[BOND93] A. Bondarenko, F. Toni, and R.A. Kowalski. An assumption-based framework for non-monotonic reasoning. *Proc. 2nd. International Workshop on Logic Programming and Non-monotonic Reasoning*, pages 171–189, 1993.

[CML98] Carlos Chesñevar, Ana Maguitman, Ronald P. Loui *Logical Models of Arguments* Submitted to ACM Computing Surveys, 1998.

[DCM98] Diego C. Martínez, *Semántica de Programas Lógicos Rebatibles*, Tesis de Licenciatura en Ciencias de la Computación, Junio de 1998.

[DCM99] Diego C. Martínez, *Alternativas para evitar falacias en la argumentación rebatible: un análisis comparativo*, Workshop de Investigadores en Ciencias de la Computación. San Juan, Junio de 1999.

[DUNG93] P. M Dung, On the Acceptability of Arguments and its Fundamental Role in Nonmotonic reasoning and Logic Programming and N-persons games, IJCAI 93 852-857, 1993.

[DUNGLP] P. M Dung, An argumentation Semantics for Logic Programs with explicit negation, ICLP 93, MIT Press.

[EREAS] *The Errors in reasoning.* http://204.38.65.2/~debate/dcbproj/support/fallacie.htm

[GCS94] Alejandro García, Carlos Chesñevar, Guillermo Simari. *The Role of Dialectics in Defeasible Argumentation* Proceedings of the XIV International Conference of the Chilean Computer Science Society. Concepción, Chile, Noviembre de 1994.

[GS99] Alejandro J. García and Guillermo R. Simari *Defeasible Argumentation for Extended Logic Programming* Submitted to the 5th International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR'99), El Paso TX, Dic. 99.

[KOWA96] Robert A. Kowalski and Francesca Toni. Abstract argumentation. *Artificial Intelligence and Law*, 4(3-4):275–296, 1996.

[LOUI97] Ronald P. Loui, Jeff Norman, Joe Altepeter, Dan Pinkard, Dan Craven, Jessica Lindsay, and Mark Foltz. Progress on room 5: A testbed for public interactive semi-formal legal argumentation. In *Proc. of the 6th. International Conference on Artifcial Intelligence and Law*, July 1997.

[PRAK] Henry Prakken y Giovanni Sartor. *Argument-based extended logic programming with defeasible priorities.* Journal of applied non classical logic 7, 25-75

[PVREE] Henry Prakken y Gerard Vreeswijk. *Logics for Defeasible Argumentation.* Survey in defeasible argumentation.

[SIM92] G. R. Simari y R. P. Loui. A Mathematical Treatment of Defeasible Reasoning and its Implementation. *Artificial Intelligence*, 53: 125–157,1992.

[VREE97] Gerard A.W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.

[YOLDER99] Brian Yoder *The Fallacy Zoo.* http://www.primenet.com/~byoder/fallazoo.htm