DERROTA Y DEFENSA EN ARGUMENTACIÓN REBATIBLE

Claudio A. Alessio
Universidad del Cuyo, Argentina

Resumen

Los sistemas argumentativos formalizan razonamiento de sentido común mediante la construcción, comparación y evaluación de argumentos a favor o en contra de ciertas afirmaciones. En tales sistemas existen dos nociones centrales: derrota y defensa. En el presente trabajo caracterizan ambas nociones y se discuten algunos desafíos para las caracterizaciones propuestas. Además, se realizan algunos comentarios al respecto.

Palabras clave: *argumentación rebatible; relaciones de derrota; relaciones de defensa.*

Recibido: 15 de octubre de 2015**. Aprobado:** 20 de marzo de 2016.

Praxis Filosófica Nueva serie, No. 45 Suplemento, julio-diciembre 2017: 25 - 53
DOI: 10.25100/pfilosófica.v0i45S.6064

ISSN (I): 0120-4688 / ISSN (D): 2389-9387

Defeat and Defense in Defeasible Argumentation

Abstract

Argumentative systems formalized common sense reasoning through construction, comparison and evaluation of arguments for or against certain conclusions. In such systems, there are two central notions: Defeat and Defense. In this paper we characterize both notions and we show and comment some challenges for proposals characterizations of defeat and defense.

Keywords: *Defeasible Reasoning; defeat relation; defense relation.*

Claudio A. Alessio. Profesor Titular de Lógica y Epistemología de la Universidad Católica de Cuyo. Becario Postdoctoral en el Instituto de Investigaciones Económicas y Sociales del Sur – CONICET. Doctor en Filosofía por la Universidad Nacional del Sur, Argentina. Sus principales áreas de trabajo y de investigación son lógica filosófica, representación del conocimiento y razonamiento y argumentación rebatible.

Dirección postal: Juramento 341 sur. Va. Las Rosas, Rawson, San Juan, Argentina (CP 5400)

Dirección electrónica: claudioalessio@uccuyo.edu.ar

DERROTA Y DEFENSA EN ARGUMENTACIÓN REBATIBLE

Claudio A. Alessio
Universidad del Cuyo, Argentina

Introducción

Los sistemas argumentativos son formalismos que modelan información tentativa y potencialmente contradictoria mediante la construcción, comparación y evaluación de argumentos a favor o en contra de ciertas afirmaciones. Diferencias específicas aparte entre los diversos sistemas argumentativos propuestos en la literatura, estos pueden ser caracterizados mediante un proceso consistente de varias etapas tal como brevemente se describe a continuación.

En la primera fase, los argumentos se construyen según ciertas reglas que deben satisfacer a partir de una base de conocimiento previamente especificada en un lenguaje formal determinado. Una vez establecido el conjunto de argumentos, puede suceder que dos o más de ellos no puedan ser simultáneamente aceptados. Cuando tal situación se da se dice que los argumentos se derrotan, o que uno derrota a otro. El conjunto de argumentos y las relaciones de derrota que entre ellos se dan, originan lo que se ha denominado '*marco argumentativo'* (Dung, 1995), constituyéndose así la segunda fase. Una vez establecido el marco, el principal objetivo para un sistema argumentativo consiste en determinar qué argumentos, de todos los construidos, pueden ser aceptados. Por ello, luego de que se tienen en consideración los argumentos y las relaciones de derrota entre ellos se procede a seleccionar aquellos argumentos que constituirán la extensión

del sistema (fase 3), i.e. el conjunto de argumentos que un agente estaría dispuesto a aceptar. Los argumentos elegidos serán los que prevalecen frente a sus rivales, entendiendo que esto los hace buenas razones para las conclusiones que sustentan, al menos en el marco argumentativo en el que se encuentran. Esta fase (fase 3) puede hacerse en base a la satisfacción de condiciones previamente especificadas, denominadas '*semánticas'*, que un conjunto de argumentos debe verificar (Dung, 1995), o mediante algún procedimiento de prueba denominado '*juegos argumentativos'* (Vreeswijk y Prakken, 2000). En ambos casos, diversas exigencias adicionales podrán pedirse a los argumentos para que estos califiquen como extensiones del sistema. Tales exigencias están regidas por criterios tolerantes, usualmente bajo teorías crédulas o criterios más estrictos bajo teorías escépticas.

Centrándose no ya en un conjunto de argumentos sino en un argumento determinado, hay acuerdo general en que el requisito mínimo que un argumento debe verificar, con vistas a que un agente pueda creer en la conclusión que sustenta, es el de ser aceptable (Dung, 1995). Un argumento será aceptable cuando, para cada argumento que lo derrota, existe al menos uno que lo defiende de ese derrotador. En términos más o menos precisos es posible decir que cuando un argumento aceptable satisface ciertos requerimientos recibe la denominación de argumento justificado. Básicamente se dirá que está justificado cuando la cadena de defensores de tal argumento descansa en un/os último/s argumento/s que no posee/n derrotador/es. Obviamente cualquier argumento que no cuenta con derrotadores es también un argumento justificado.

Ya sea mediante las *semánticas* o mediante *juegos*, la selección de los argumentos es realizada en base a su interacción en el marco argumentativo al que pertenecen, i.e. a las relaciones de derrota que entre los argumentos se dan. La noción de aceptabilidad permite notar que un argumento, aunque se encuentre derrotado por otro, no necesariamente significa que tal argumento debe rechazarse. Puede suceder que un argumento *'b'*, que oficia como derrotador de otro *'a'*, esté a su vez derrotado por un argumento *'c'*, de manera que *'a'*, se dice en sistemas argumentativos, puede verse restablecido. Por ello, para determinar el estado final de un argumento será necesario conocer todas las interacciones entre los argumentos, incluida la del restablecimiento. El restablecimiento puede ser ilustrado con el siguiente ejemplo.

Ejemplo 1.1: Tom Grabit
Supóngase que el profesor X tiene razones para creer 'a': El profesor X ha visto a Tom Grabit robar un libro en la biblioteca por lo que puede concluir que Tom Grabit robó un libro de la biblioteca. Ahora supóngase que el profesor X cuenta con un derrotador 'b' de 'a': la Señora Grabit dice que Tom está a miles de kilómetros de distancia y su hermano gemelo, que es cleptómano, estaba en la biblioteca el día en que el Profesor X supuestamente vio a Tom. Ahora bien, si luego el profesor X se entera, por su psiquiatra, que la Señora Grabit es una mentirosa compulsiva y desquiciada, y que el hermano gemelo de Tom es un invento de su mente, entonces ha adquirido un derrotador 'c' para 'b'.

En el ejemplo 1.1 originalmente propuesto por Lehrer y Paxson (1969) es posible constatar que *'b'* hace que *'a'* sea considerado injustificado, pero *'c'* restaura la justificación original de *'a'*. La intuición subyacente en el restablecimiento consiste en que un argumento prevalecerá frente a sus adversarios (contará como parte de una extensión de un marco argumentativo, o tendrá una prueba en un juego argumentativo) cuando todos sus posibles derrotadores estén a su vez derrotados. Una discusión sobre la validez del restablecimiento puede encontrarse en (Horty, 2001; Prakken, 2002; Bodanza y Alessio, 2014; Alessio, 2015)

Además del restablecimiento, otra propiedad es exigida en el marco de los sistemas argumentativos, denominada como principio de composicionalidad (Vreeswijk, 1997), para considerar si, en última instancia, un argumento puede considerarse justificado. Este principio exige que si un argumento *'a'* está justificado entonces *'a'* se encuentre sustentado por subargumentos justificados. Por ejemplo:

Ejemplo 1.2
a. Dado que Nixon es pacifista, puesto que por lo general los cuáqueros son pacifistas y que Nixon es cuáquero, y teniendo en cuenta que los pacifistas no usan armas, se puede concluir que Nixon no usa armas.
b. Teniendo en cuenta que por lo general los republicanos no son pacifistas y que Nixon es republicano se puede concluir que Nixon no es pacifista.

Supóngase que por alguna razón la información con respecto a ser republicano tiene prevalencia sobre el ser cuáquero, de modo que el argumento *'b'* derrota a un subargumento del argumento *'a',* el que sustenta que Nixon es pacifista. Ahora bien ¿se podría decir que el argumento *'a'*

está justificado? Claramente, no. La razón para tal respuesta se debe a que una información clave para creer en que Nixon no usa armas, ha sido desacreditada por creer que Nixon es republicano, y en consecuencia no verifica el principio de composicionalidad nombrado anteriormente.

Atendiendo a que los sistemas argumentativos son formalismos que modelan información tentativa y potencialmente contradictoria mediante la construcción, comparación y evaluación de argumentos a favor o en contra de ciertas afirmaciones, el objetivo central del presente trabajo será realizar una caracterización de distintas relaciones que pueden darse entre argumentos en el contexto los sistemas nombrados. En particular en lo que respecta a las relaciones negativas o de rivalidad, como la de derrota y el conflicto, y también de las relaciones positivas o de alianza entre argumentos como la de defensa. Además de una caracterización de tales relaciones se harán algunos comentarios y se señalarán algunos desafíos o dificultades.

El trabajo se organiza en tres secciones. En la primera se exponen diversas relaciones de rivalidad entre argumentos. La segunda se centra en intentar capturar la noción de defensa. La tercera sección retomará los conceptos expuestos y se discute la relevancia filosófica de los mismos. Finalmente se concluye.

Derrota entre argumentos

Los sistemas argumentativos gravitan en torno a la noción de argumento en tanto prueba tentativa de la conclusión que sustentan. Los sistemas propuestos en la literatura (e.g. Simari y Loui, 1992; Pollock, 1995; Prakken, 1993; Vreeswijk, 1993) establecen diversas maneras de definir un argumento, pero es claro que un argumento está constituido por un conjunto de premisas (que ofician de razones para una conclusión), una conclusión y una relación de justificación entre las razones y las conclusiones.

Pollock (1987) sostiene que existen dos tipos de argumentos, los estrictos y los rebatibles. Los argumentos estrictos son definidos en base a razones conclusivas, i.e. razones que implican lógicamente una conclusión. Todos los razonamientos deductivos forman parte de este tipo de argumentos. Los argumentos rebatibles, por su parte, son construidos a partir de razones *prima facie.* Una razón *prima facie* brinda un sustento razonable pero tentativo para una conclusión determinada. Información adicional puede invalidar ese sustento, como en los ejemplos 1.1 y 1.2. Otro ejemplo típico en argumentación rebatible es el de Tweety (ejemplo 2.1). Este ejemplo muestra que puede obtenerse la conclusión tentativa *"Tweety vuela"* en base a razones *prima facie* "*Tweety es ave y el hecho de ser ave es una buena razón para creer que vuela".* Ahora bien, la información *"Tweety*

es pingüino" invalida tal argumento, ya que los pingüinos son aves que no vuelan, obviamente tal derrota es tentativa a su vez, puesto que *Tweety* podría ser un pingüino excepcional.

Además de ejemplos como el de Tweety, es posible identificar una gran variedad de razonamientos rebatibles. Pollock (1987) dice que en general todos los argumentos que apelan a información brindada por la percepción, la memoria, el uso de inferencias estadísticas o argumentos construidos a partir de testimonios permiten la construcción de argumentos que dan buenas razones para la conclusión que sustentan, sin embargo, éstas no quedan establecidas de manera firme pudiendo ser revisadas en presencia de argumentos mejores.

La característica esencial de la derrotabilidad de un argumento está dada por la posibilidad de encontrar contraargumentos. Ahora bien, la existencia de objeciones contra ciertos argumentos no significa que tales argumentos no sean buenas razones para las conclusiones que sustentan, puesto que el contraargumento puede ser un *'mal'* contraargumento.

Por lo dicho es menester aclarar cuándo un contraargumento es un *'buen'* contraargumento y cuándo será un '*mal'* contraargumento, al menos en lo que respecta a los sistemas argumentativos. Para ello será necesario aclarar el significado de dos conceptos clave: derrota (*defeat*) y defensa (*defense*).

El objetivo de la presente sección será responder a la siguiente pregunta: *¿Cuándo un argumento derrota a otro?* Supóngase que a partir de un conjunto de información inicial pueden construirse los siguientes argumentos:

Ejemplo 2.1: Tweety el pingüino
a. *A partir del dato de que Tweety es ave y dado que por lo general las aves vuelan es posible concluir que Tweety vuela.*
b. *A partir del dato de que Tweety es pingüino y dado que por lo general los pingüinos no vuelan es posible concluir que Tweety no vuela.*

En el ejemplo es fácil advertir que *'a'* y *'b'* no pueden ser conjuntamente aceptados de un modo racional ya que se aceptaría que *Tweety vuela* y *que no lo hace* simultáneamente. Cuando dos argumentos no puedan simultáneamente ser aceptados, se dice que tales argumentos están en conflicto o en interferencia. El conflicto es una condición necesaria para la derrota, pero no suficiente tal como se verá más adelante. A continuación, se pretenderá precisar el concepto de conflicto.

En la literatura pueden encontrarse tres tipos de conflictos (o ataques). El conflicto conclusión-conclusión o conflicto por rebatimiento (*rebutting attack*); el conflicto conclusión-premisa o ataque a premisas (*undermining attack)* y el conflicto conclusión-inferencia o conflicto por socavamiento (*undercutting attack)*.

El primer tipo de conflicto: conclusión-conclusión, se da cuando las conclusiones de los argumentos en consideración son contradictorias, como en el ejemplo 2.1. El siguiente ejemplo, también es clásico en la literatura, conocido como "*Diamante de Nixon*" ilustrará este tipo de conflicto.

Ejemplo 2.2: Diamante de Nixon
a. *Teniendo en cuenta que por lo general los cuáqueros son pacifistas y que Nixon es cuáquero se puede concluir que Nixon es pacifista.*
b. *Teniendo en cuenta que por lo general los republicanos no son pacifistas y que Nixon es republicano se puede concluir que Nixon no es pacifista.*

'a' y 'b' se dicen en conflicto por rebatimiento porque la conclusión de un argumento niega la conclusión del otro y viceversa. Ahora bien, no necesariamente las conclusiones deben ser explícitamente contradictorias. Puede haber conflicto por rebatimiento y las conclusiones de ambos argumentos se refieran a cuestiones diferentes. La razón por la que se dirá que están en conflicto por rebatimiento será porque la *conjunción de sus conclusiones implica contradicción.*

Ahora considérese el ejemplo 2.3.

Ejemplo 2.3
a. Juan nació en el país X y por lo general los nacidos en el país X hablan el idioma P. Si Juan habla el idioma P entonces cuando viaje al país H (que también hablan el idioma P) Juan no tendrá problemas de comunicación.
b. Juan nació en la aldea F del país X y por lo general los nacidos en la aldea F no hablan el idioma P. Por lo tanto, Juan no habla el idioma P.

En el ejemplo 2.3, la conclusión del argumento *'b'*: '*Juan no habla el idioma P'*, niega un paso en las premisas del argumento *'a'*, a saber, '*Juan habla el idioma P*'. Como se advierte la contradicción no emerge a partir de las conclusiones de ambos argumentos sino entre la conclusión de uno y una premisa del otro. Este tipo de conflicto se denomina conflicto conclusión-premisa o ataque a premisas y se da cuando la conclusión de un argumento

niega una premisa de otro argumento, o lo que es lo mismo, cuando niega una conclusión sustentada por un sub-argumento del argumento atacado (como en el ejemplo 2.3). Otro caso similar puede ser ilustrado en el ejemplo 2.4. En tal ejemplo se niega un supuesto al que apela uno de los argumentos.

Ejemplo 2.4
a. Tweety es Ave (P) y no es probable que Tweety sea pingüino (-R) por lo tanto, Tweety vuela (T).
b. Los medios usados para la observación de Tweety fueron precisos y fiables (Q), por lo tanto, se llega a la conclusión de que Tweety es pingüino (R).

En el ejemplo 2.4 se ve que una premisa del argumento *'a'*, esto es, *-R* está en conflicto con la conclusión *R* de *'b'*. Como puede fácilmente advertirse, es un caso de conflicto conclusión-premisa (Prakken & Sartor, 1995).

También se debe advertir que el conflicto conclusión-premisa aparecerá, no cuando haya, necesariamente, explicita información contradictora, sino *cuando la conjunción de la conclusión de un argumento y una premisa de otro implique contradicción.* Esto abre el abanico a una multitud de casos en los que puede darse el conflicto conclusión-premisa. Por caso, imagine un argumento tal que dos premisas implican una proposición contradictoria con una proposición implicada por la conclusión del otro argumento.

Finalmente, el último tipo de conflicto detectado en la literatura es el conocido como conflicto por socavamiento o conflicto conclusión-inferencia. Tal conflicto fue propuesto inicialmente por Pollock en (1987). Se da cuando la conclusión de un argumento niega la conexión entre premisas y conclusión del otro o establece la falta de sustento que las premisas ofrecen a la conclusión. El ejemplo canónico de ataque por socavamiento es el siguiente.

Ejemplo 2.5
a. La mesa que esta ante mí, parece color rojo (P), por lo tanto, es posible concluir que la mesa es roja (T)
b. La mesa que está ante mí parece color rojo (P), porque se encuentra iluminada por una luz roja (Q), esto me hace dudar acerca de la conexión entre las premisas y conclusión de A.

A diferencia de los ejemplos considerados hasta aquí puede advertirse que el argumento *'b'* no niega la conclusión del argumento *'a'* ni tampoco alguna de las premisas, de hecho, no es definido en base a la implicación de contradicción, sino que es un argumento que señala que las premisas del

argumento en cuestión no alcanzan, dada la información disponible, para sustentar adecuadamente la conclusión.

Esquemáticamente, los conflictos considerados hasta aquí pueden representarse en la Figura 2.1. Las flechas punteadas, representan relaciones de inferencia, las flechas sólidas representan relaciones de conflicto y '-(a>b)' representa la negación de la inferencia de a a b:

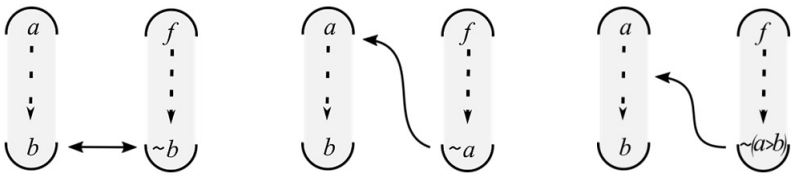

Figura 2.1 Conflictos: rebatimiento, ataque a premisas, socavamiento

Aunque la única relación de conflicto simétrica es la de rebatimiento, es posible identificar conflictos mutuos de ataque a premisas (*crossover attack*).

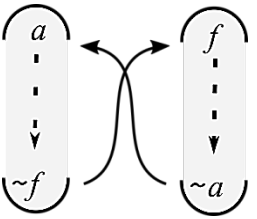

Figura 2.2 Ataque a premisas mutuo

Ejemplo 2.6
a. *Dado que Uki es ave, Uki anida en árbol y dado que anida en árbol y es ave; entonces Uki vuela.*
b. *Dado que Uki es un ñandú, Uki no vuela y dado que es ave y no vuela, Uki no anida en árbol*

El ejemplo 2.6, debido a Chesñevar (1996), ilustra la idea de ataque cruzado o ataque a premisas mutuo. La conclusión del argumento *'b': 'Uki no anida en árbol'* niega una *sub-conclusión* del argumento *'a'* ('*Uki anida en árbol'*). La conclusión de *'a'* niega una *sub-conclusión* de *'b'*.

El socavamiento también tiene una versión mutua o cruzada. El ejemplo fue propuesto por Pollock (2001) y es presentado en el ejemplo 2.7.

Ejemplo 2.7
a. Juan dice que Pedro no es digno de confianza (‘q’). Por lo tanto, es posible concluir que Pedro no es digno de confianza (‘r’).
b. Pedro dice que Juan no es digno de confianza (‘s’). Por lo tanto, es posible concluir que Juan no es digno de confianza (‘t’).

Es claro que *‘r’* es una razón para creer que *‘s’* no es una buena razón para ‘*t*’. Al mismo tiempo, *‘t’* es una razón para creer que *‘q’* no es una buena razón para *‘r’*.

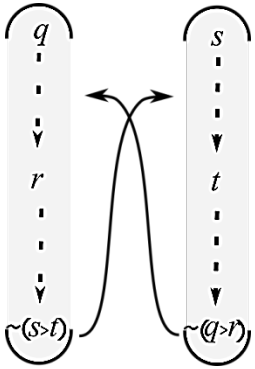

Figura 2. 3 Socavamiento mutuo

La noción de conflicto no dice nada acerca del éxito de los argumentos rivales, simplemente señala la imposibilidad de su aceptación conjunta. Para determinar cuál de ellos prevalece es necesario algún mecanismo que permita decidir cuál (o ninguno) debe aceptarse. Esta necesidad lleva a la otra noción básica en sistemas argumentativos, la noción de derrota (*defeat*). La misión de esta consiste en determinar el éxito de un conflicto entre pares de argumentos. Intuitivamente se entiende que un argumento derrota a otro cuando ambos no pueden ser aceptados conjuntamente, pero uno de ellos es mejor, o dicho de una forma más amplia, cuando uno de ellos no es peor que el otro.

En general, la relación de derrota es definida a partir de la relación de conflicto y de un orden previamente establecido entre los argumentos. El orden es realizado mediante una relación de preferencia. Dependiendo del tipo de información modelada es posible identificar diversos tipos de preferencia para resolver los conflictos, por ejemplo, prioridades de leyes en un sistema legal, deseos o valores en el razonamiento práctico, confiabilidad en el razonamiento epistémico, o especificidad en razonamiento default. Sin embargo, Modgil y Prakken (2013) advierten que existen derrotas dependientes de la relación de preferencia y derrotas que no lo son. Esto será

tenido en cuenta a la hora de presentar los diversos tipos de derrota, que al igual que las relaciones de conflicto, son tres las propuestas en la literatura: derrota por rebatimiento (*rebutting defeater*), derrota por socavamiento (*undercutting defeater*) y derrota de premisas (*undermining defeater*).

La definición de derrota por rebatimiento, empleada aquí, fue propuesta por Prakken y Sartor (1996) pero vale la pena aclarar que existen otras.

Definición 2.1: Derrota por rebatimiento

Sean 'a' y 'b' dos argumentos. Si la conclusión de 'a' y la conclusión de 'b' implican contradicción y 'b' no es preferido a 'a' se dice que 'a' ***rebate*** *a 'b'.*

[Prakken & Sartor, 1996b]

Nótese que la definición de rebatimiento apela a la noción de preferencia, esto es así porque, en caso de estar frente a un conflicto por rebatimiento, algo debe hacer que se incline la balanza para un lado u otro. Obviamente, si se está frente al caso planteado por el diamante de Nixon, al no haber una preferencia explicita por alguno de los dos, es útil considerar que uno derrotará al otro cuando este no sea preferido al primero.

Una cuestión que merece la pena advertir aquí es el hecho de que casos como el ejemplo de Tweety (ejemplo 2.1), tradicionalmente han sido entendidos como rebatimiento asimétrico (*asimetric rebutting defeater*), y se ha afirmado que la preferencia a la que se apela es la de especificidad. Ahora bien, lo anterior es discutible puesto que parece más correcto entenderlo como un caso de socavamiento (Pollock, 2001; Walton, 2011). Los casos de derrota por especificidad, como el de *Tweety*, suponen un interesante desafío para enriquecer lo que se entiende por derrota, Modgil y Prakken (2013) afirman que, si el sistema que modela tales casos tiene el suficiente poder expresivo, en tal ejemplo se da una derrota por socavamiento. Tal vez sea esa la razón de por qué ejemplos que apelan a la especificidad como criterio para dirimir conflictos lleva a problemas de representación como lo destaca Horty (2001). Si esta idea es plausible, será necesario investigar si el ejemplo de Tweety debe modelarse como un caso de rebatimiento o de socavamiento, y una vez resuelta la disyunción, decir cómo hacerlo en un sistema específico que actualmente lo modela como rebatimiento tales como en (Prakken y Sartor, 1996) o en (García y Simari, 2004).

Los ejemplos 2.3 y 2.4 ilustran los dos casos de derrota a premisas, uno dependiente de la relación de preferencia mientras que el otro no lo es. La diferencia estará dada por la naturaleza de la premisa que es atacada. Las preferencias son necesarias para resolver conflictos, excepto cuando la

premisa establece alguna hipótesis que apela a la ausencia de evidencia de lo contrario, por ejemplo, negación por falla en programación lógica como en el ejemplo 2.4. Pero el siguiente ejemplo (2.8) muestra la necesidad de apelar a un criterio de preferencia frente a cierto tipo de casos.

Ejemplo 2.8

a. Juan, especialista en violines, dice que el violín en cuestión es caro porque es un Stradivarius. Luego, se puede concluir que el violín es caro.

b. Luis, el hijo de Pedro, de tres años de edad dice que el violín no es un Stradivarius. Luego, se puede concluir que el violín no es un Stradivarius.

Si la derrota a premisas no involucrara preferencias, entonces, el argumento *'b'* derrotaría al argumento *'a'* pero no parece tener sentido aquí. Si se establece un criterio de confiabilidad (o autoridad), la sentencia '*no es un Stradivarius'* no derrota la premisa: *'Juan dice: El violín es un Stradivarius'*, porque el hijo de Pedro no es confiable en el tema en cuestión. tal como intuitivamente se entiende el ejemplo.

La derrota a premisas dependiente de preferencias puede ser definida como un rebatimiento (indirecto) entre un subargumento propio de *'a'* y el argumento *'b'* de la siguiente manera:

Definición 2.2: Derrota por rebatimiento indirecto

Sean 'a' y 'b' dos argumentos. Si la conclusión de 'b' y la conclusión de 'a' (donde 'a*' es un subargumento propio de 'a') implican contradicción y 'a*' no es preferido a 'b', 'b' derrota por* ***rebatimiento indirecto*** *a 'a'.*

[Prakken & Sartor, 1996b]

Atendiendo a la definición 2.2, en el ejemplo 2.8 el argumento *'b'* no es capaz de derrotar el subargumento *'Juan, especialista en violines dice que el violín es un Stradivarius'*, por lo tanto, es razonable concluir que el violín es caro (por ser un Stradivarius).

El ejemplo 2.4, un caso de derrota a premisas, no es dependiente de la noción de preferencia. Por tal motivo es menester distinguirla de la anterior. Con vista a evitar confusiones se llamará a este tipo de derrota: *derrota a hipótesis.*

Definición 2.3: Derrota a hipótesis
Sean 'a' y 'b' dos argumentos. Si la conclusión de 'a' niega una hipótesis en 'b', 'a' ***derrota*** *a 'b'.* [Prakken & Sartor, 1996b]

Finalmente, resta por considerar la derrota por socavamiento. El socavamiento no exige la explicitación de un criterio preferencia para ser definido. El mismo puede ser definido, siguiendo a Pollock (1987) de la siguiente manera.

Definición 2.4: Derrota por socavamiento
Sean 'a' y 'b' dos argumentos. Si la conclusión de 'a' niega la relación de inferencia en 'b', 'a' ***derrota por socavamiento*** *a 'b'.* [Pollock, 1987]

Hasta aquí se han presentado ejemplos y se ha visto una tipología de conflicto y derrota. Se han brindado algunas precisiones y se han advertido ciertas cuestiones que en la literatura no parecen aun resueltas como la derrota por especificidad. Ahora es tiempo de dar una respuesta más clara sobre *qué significa que un argumento derrota a otro* y *cuándo puede decirse que un contraargumento es un 'buen' ('mal') contraargumento.*

Según lo expuesto, la derrota entre argumentos se basa esencialmente en la imposibilidad de aceptación conjunta y no únicamente en la existencia de contradicciones. Esto es interesante puesto que ello implica que puede haber imposibilidad de aceptación conjunta pero no inconsistencia. Además, es útil puesto que pueden definirse sistemas con diferentes propósitos, claro que un sistema para la resolución de argumentos conflictivos, como es usual en sistemas argumentativos, es lo natural, sin embargo, si se quisiese que el sistema arrojara un único argumento para cada conclusión, por ejemplo, aunque sean inconsistentes esto podría hacerse sin dificultad atendiendo a esta amplia concepción del conflicto.

La derrota entre aquellos argumentos, cuya relación de conflicto exige preferencias para su resolución, puede ser definida como: un argumento derrota a otro cuando es imposible aceptarlos conjuntamente y uno no es mejor que el otro.

Pero es importante advertir que la definición anterior no es tan general como para cubrir a todas las relaciones de derrota. Supóngase que '*a'* socava a '*b*'. Es claro que '*a*' y '*b*' no son aceptables conjuntamente y que '*a*' no es preferido (en el sentido técnico) a '*b*' (supóngase porque son incomparables

según las relaciones de preferencia definidas en el sistema), entonces *'a'* debería derrotar también a *'b',* lo cual no tiene sentido.

Otro comentario que podría hacerse al respecto de la derrota es sobre el efecto de tal relación en un argumento, i.e. sobre el estado de derrotado de un argumento. Al respecto cabe señalar que tal estado es, en general, retractable. *'a'* puede derrotar a *'b'* en el tiempo t con el conocimiento K. Ahora bien, en el tiempo t_2 con el conocimiento K_2, *'a'* puede no derrotar a *'b'* porque, por ejemplo, la imposibilidad de aceptación conjunta se ha esfumado, o siguen siendo incompatibles pero las preferencias han cambiado, o las preferencias se mantienen y se mantiene la imposibilidad de aceptación conjunta pero la derrota no es efectiva por haber sido neutralizado su poder derrotador por efecto de un derrotador del derrotador.

Con respecto a la pregunta sobre cuándo un contraargumento puede considerarse un *buen contraargumento* la respuesta parece sencilla, *un contraargumento puede considerarse un buen contraargumento cuando este es exitoso, en caso contrario podría llamarse mal contraargumento.* En Sudduth (2008) se discuten las diferencias entre derrotadores aparentes (*misleading defeaters*) y derrotadores genuinos (*genuine defeaters*), un derrotador aparente es aquel que en el estado actual del conocimiento derrota a un argumento determinado pero tal derrotador no es efectivo porque existe un derrotador genuino de tal derrotador. Un derrotador genuino es aquel que efectivamente derrota y, o bien no es derrotado, o está defendido por un argumento defendido o no derrotado. Para una mejor comprensión de la derrota será necesario aclarar el significado de defensa que se hará en la sección siguiente.

Defensa entre argumentos

Los argumentos rebatibles son pruebas tentativas para la conclusión que sustentan y pueden ser abandonados frente a argumentos mejores. La rebatibilidad de un argumento expresa una especie de vulnerabilidad epistémica que puede verse actualizada si se cuenta con razones que nieguen o reduzcan la razonabilidad del argumento en cuestión. A pesar de la vulnerabilidad epistémica de un argumento rebatible, este puede gozar de un estado epistémico positivo i.e. un argumento en el que es razonable creer dado el conocimiento disponible. Ahora bien, cuando es *derrotado*, tal argumento pierde ese estado. Pero esto merece una atenta aclaración. Para que un argumento pierda su estado, no solo debe ser derrotado, sino que su derrotador tiene que poseer un estado epistémico positivo, puesto que de lo contrario no parece razonable decir que es un derrotador que derrota en

sentido *genuino* y sería más adecuado llamarlo como *derrotador aparente*. Esto puede ser ilustrado recurriendo nuevamente al ejemplo 1.1.

Supóngase que el profesor Conforti cree en lo siguiente: *Tom Grabit robó un libro en la biblioteca.* Esta conclusión se sustenta en la siguiente observación del Profesor: *he visto a un individuo que se parece a Tom Grabit robar un libro de la biblioteca.* Con tal información disponible Conforti puede razonablemente creer que Grabit robó un libro. Ahora bien, supóngase que Conforti se encuentra con la Señora Grabit quien le dice: *Tom está a miles de kilómetros de distancia y su hermano gemelo, que es cleptómano, estaba en la biblioteca el día en que Usted cree haber visto a Tom.* Esta información, más el hecho de que Conforti no tiene razones para dudar de lo que dice la Señora Grabit, le da una *fuerte razón* para creer que su creencia sobre Tom no es razonable.

Hasta aquí, un argumento rebatible ha actualizado su vulnerabilidad epistémica y el estado de argumento razonable que ostentaba se ha perdido. Ahora bien, supóngase que Conforti se encuentra por casualidad con el psiquiatra de la Señora Grabit, amigo de Conforti, que tras contarle lo sucedido, le dice*: la Señora Grabit es una mentirosa compulsiva y desquiciada, y el hermano gemelo de Tom es un invento de su mente.* Cuando tal información es adquirida por Conforti, su creencia original recupera su estado inicial.

Como se puede notar en el ejemplo de Tom Grabit, el argumento rival a la creencia original no era más que un *derrotador aparente*, al menos en el estado actual de conocimiento. Lo relatado anteriormente permite ilustrar lo que en argumentación rebatible se denomina *defensa* o *restablecimiento* de argumentos (*argument reinstatement*). En esta sección se pretenderá brindar una clarificación de tal concepto. Una primera aproximación parece sugerir que la defensa puede ser capturada con la definición 3.1.

Definición 3.1: Defensa General
*Sean 'a' y 'b' argumentos. Se dice que 'a' **defiende** a 'b', cuando 'a' derrota a un derrotador de 'b'.*

La definición 3.1, en principio parece adecuada. Lo que lleva a creer esto es el hecho de que cuando un argumento derrota al derrotador de otro, parecería que tal argumento ejerce un rol de defensor en el sentido intuitivo de la expresión. A continuación, se analizará si es correcta tal como a primera vista parece.

Con vistas a estudiar fácilmente una gran variedad de casos, los argumentos serán expresados por letras minúsculas y las relaciones de

derrota por flechas sólidas. Esto conformará *grafos argumentativos* (i.e. un conjunto de nodos, los argumentos, y un conjunto de aristas, las relaciones de derrota). Cuando sea necesario, las relaciones de defensa serán representados mediante flechas punteadas.

La representación '$a \rightarrow b$' debe interpretarse como "*el argumento 'a' derrota al argumento 'b'*". '$a \leftrightarrow b$' debe leerse como "'*a' y 'b' se derrotan mutuamente*". '$a \dashrightarrow b$' se leerá como *"'a' defiende a 'b'"* pudiendo darse el caso simétrico. El ejemplo 1.1 puede representarse como en la Figura 3.1 (I) donde *'b'* derrota a *'a' y 'c'* derrota a *'b'* y atendiendo a la definición 3.1, *'c'* debe considerarse un defensor de *'a'*.

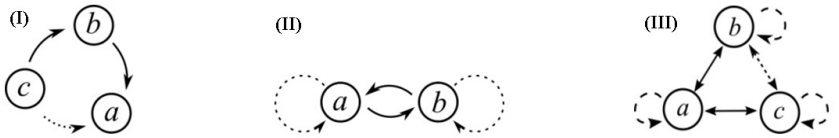

Figura 3. 1

El caso de Tom Grabit, o mejor, la estructura representada en la figura 3.1 será llamada a partir de aquí como restablecimiento (o defensa) canónico (canónica) para facilitar su referencia a lo largo del artículo. El ejemplo 2.3 por su parte se podría representar como en la Figura 3.1 (II) ya que es claro que el argumento '*a'* defiende a *'a'* de *'b'*. Lo mismo puede decirse de '*b'*. La situación ilustrada por el ejemplo 2.3 se denomina como *autodefensa*.

Ahora supóngase una situación donde hay tres argumentos: *'a', 'b'* y *'c'*. Adicionalmente supóngase que '*b*' y '*a*' se derrotan mutuamente, lo mismo entre '*c*' y '*a*'. En la Figura 3.1 (III) se ilustra la situación. Lo mismo que sucede en el ejemplo 2.3 aquí sucede, cada uno de los argumentos se defiende frente a los ataques de otros argumentos, pero, además, hay aquí un comportamiento distinto, comportamiento que podría ser llamado *defensa mutua*. Esto es así entre *'b'* y *'c'*, puesto que '*b*' defiende a '*c*' de '*a'* y '*c*' defiende a '*b*' de '*a*'.

Otros casos interesantes de defensa son los que pueden verse representados en las Figuras 3.2 (I) y 3.2 (II) Nótese que en la representación de la figura 3.2 (I), el estado del argumento '*c',* i.e. si el argumento es aceptado o no por el agente que lo propone, depende del estado de *'a'* y de '*d'* simultáneamente. Si alguno de los dos es abandonado, o si el oponente cuenta con un derrotador para cualquiera, el estado de *'c'* cambiaría.

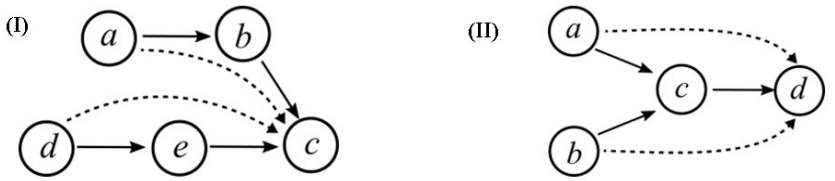

Figura 3. 2

Por su parte, en la Figura 3.2 (II) *'d' depende* de '*a'* o *'b'*. Si en el transcurso de un debate, por ejemplo, el agente que defiende '*d'* recibe una crítica en lo que respecta al argumento '*a'* eso no significará que deba retractarse de '*d'* puesto que el derrotador de '*d',* i.e. *'c',* se encuentra *también* derrotado por *'b'*.

Retomando el tema central de esta sección, se puede decir que hasta aquí la definición sugerida de defensa parece intuitiva y adecuada. Sin embargo, a continuación, se considerarán algunos ejemplos que permiten ilustrar el hecho de que lleva a resultados un poco extraños y ponen en duda que la definición 3.1 sea una correcta caracterización de la noción de defensa intuitiva. Suponga el caso de un argumento que se *autoderrota*. Llamativamente la definición de defensa general lleva a aceptar que *un argumento que se autoderrota, es también un argumento autodefendido,* lo que no parece tener sentido (Figura 3.3 (I)).

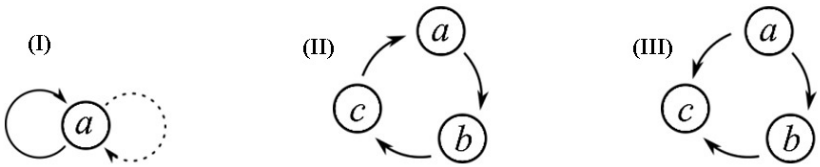

Figura 3.3

Por otro lado, considérese la situación ilustrada en la Figura 3.3 (II). En ese caso, ¿Será correcto decir que el argumento *'a' defiende* a '*c'?* Siguiendo la definición propuesta la respuesta parece afirmativa. De modo que otra consecuencia extraña de la definición 3.1 es obtenida: *un argumento puede defender a su derrotador*. En la Figura 3.3 (III), ¿podrá decirse que el argumento *'a'* defiende a '*c'?* Intuitivamente no parecería adecuado afirmar que '*a'* es un defensor de '*c'*, sin embargo, si se *asume* la definición 3.1 se *debe* aceptar que *un argumento 'a' puede defender y derrota a otro 'b' al mismo tiempo.*

La Figura 3.4 (I), un poco más compleja, exige pensar si *'b'* es un defensor de *'e'*. En principio parece no haber dudas, ahora bien, la definición

3.1 lleva a aceptar que *habrá argumentos que defiendan a ciertos argumentos y que simultáneamente defenderán a argumentos rivales de tales argumentos.*

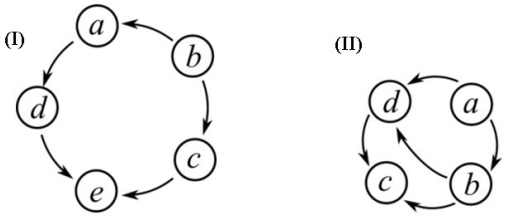

Figura 3.4

La Figura 3.4 (II) plantea ya una situación más compleja y tal vez más extrema, ¿*'a'* defiende a '*c'*? y ¿ *'a'* defiende a *'d'*? A primera vista '*a*' defiende a '*c*' puesto que derrota a todos los derrotadores de este argumento, sin embargo, una mirada más profunda permite evidenciar que '*a*' es un defensor de un agresor de '*c*', a saber, '*d*'. En caso de no estar dispuestos a aceptar como defensa casos como los considerados, algunos refinamientos pueden ser realizados a la definición 3.1. Se podría exigir, en una primera instancia, que los argumentos involucrados en una defensa no sean rivales. El comportamiento frente a los ejemplos de las Figuras 3.4 (I) y 3.4 (II) podrían ser tolerados, pero definitivamente no parece correcto sostener que argumentos inconsistentes mantengan una relación de defensa.

Definición 3.2: Defensa consistente
Se dice que 'a' es un ***argumento defensor**** *de 'b', notado como a⇢b, cuando 'a' derrota a un derrotador de 'b' y ni 'a' derrota 'b' ni 'b' derrota 'a'.*

Atendiendo a este nuevo concepto de defensa, denominada *defensa consistente*, será interesante evaluar si puede modelar correctamente lo que modela la defensa general y si filtra los casos problemáticos. Al respecto, el ejemplo canónico sigue siendo modelado de la misma manera, puesto que '*c*' defiende a '*a*' también en este sentido. La autodefensa en el ejemplo de Nixon también sigue siendo modelada correctamente. Una innovación ocurre si un argumento se derrota a sí mismo, *la definición 3.2 no permite la autodefensa* en casos de *autoderrota*. A continuación, se muestra una comparación entre el comportamiento de la defensa general (definición 3.1) y la defensa consistente (definición 3.2).

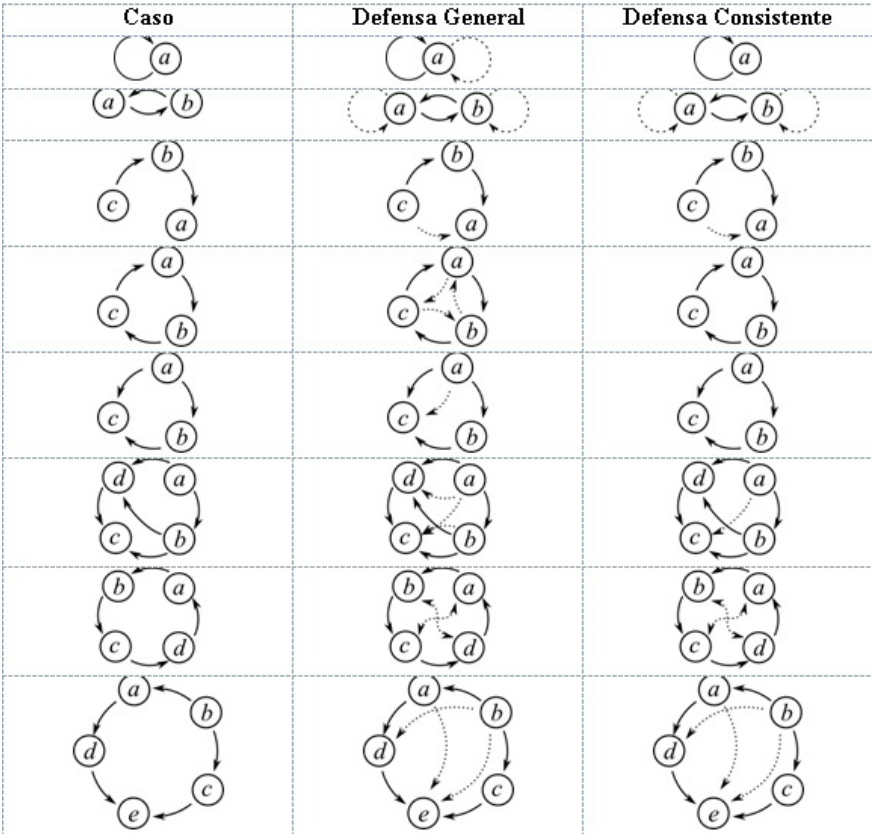

Figura 3.5 Defensa consistente

Como puede observarse el comportamiento es considerablemente diferente en la propuesta de la definición 3.2 comparado con la anterior.

Hasta aquí se ha intentado responder a la pregunta: ¿Cuándo un argumento puede considerarse defensor de otro? A primera vista parecía que la derrota de un derrotador constituía defensa, pero un análisis más específico facilitó la advertencia de que tal idea permitía aceptar que un argumento se considere defensor de otro cuando, y al mismo tiempo, derrote al defendido. Ello motivó un segundo intento de caracterización. El resultado, exigir, además de la derrota de derrotadores, que defensor y defendido sean libres de conflicto. Tal noción parece una buena candidata para capturar la idea intuitiva de defensa. Obviamente, también *adolece del resultado problemático* ilustrado en la Figura 3.4 (II) por eso cabe la pregunta: ¿puede decirse que un argumento '*a*' que derrota al derrotador '*b*' de otro argumento '*c*' ('*a*' y '*c*' son consistentes) pero que al mismo tiempo

derrota a un derrotador '*d*' de un rival '*e*' de tal argumento ('*e*' derrota a '*c*'), es un argumento defensor de '*c*'? Este tipo de situaciones serán llamadas derrotas indirectas.

Desde un punto de vista práctico, no parece razonable emplear como defensor de un argumento *'c'* a un argumento *'a'* que neutraliza la fuerza inhibidora de *otro* defensor de *'c'*. Por ejemplo, supóngase que un proponente quiere dar pruebas a favor de *'c',* el oponente derrota a '*c*' con '*d*', el proponente *defiende* a '*c*' con '*e*'. Luego, el oponente derrota a '*c*' con '*b*' y el proponente para *defender* a *'c'* de *'b'* emplea *'a*' pero resulta que '*a*' también es un derrotador de '*e*'. De modo que no parece una estrategia adecuada. Esto sugiere que la noción de defensa debe ser modificada e incorporar de algún modo la restricción de las derrotas indirectas. A pesar de ello, esta tarea será postergada para otra ocasión, sirva al menos esta sección para dejar sentado el problema para una resolución futura. Resolución que deberá atender también a lo siguiente.

Además de lo considerado sobre las derrotas indirectas un aspecto resta aún por discutir: *cuando se habla de defender a un argumento ¿se quiere decir que el defensor tiene éxito en su pretensión de defensa?* La pregunta tiene sentido porque tanto la definición 3.1 como la 3.2 consideran que hay *defensa* cuando el caso puede catalogarse como *defensa no exitosa*. Considérese los siguientes ejemplos que ilustrarán lo que se está diciendo. Supóngase que '*a*' derrota a '*b*', '*b*' derrota a '*c*' y '*c*' derrota a '*d*'. Bajo la definición 3.2, *'a' es un defensor de 'c'* puesto que *'a' derrota a un derrotador de 'c'.* Por su parte, nótese también que *'b' es un defensor de 'd'* ya que *'b' derrota a un derrotador de 'd'.* Ahora bien, ¿hay alguna diferencia entre ambas situaciones? Tal como es obvio, en ambos casos, '*a*' y '*b*' son defensores (de '*c*' y '*d*' respectivamente) en el sentido de la definición 3.2. Ahora bien, mientras que '*d*' es defendido por un argumento derrotado, '*c*' es defendido por un argumento no derrotado.

Ahora suponga que '*v*' derrota a *'x', 'x'* derrota a *'y', 'y'* derrota a *'z', 'z'* derrota a *'t'*. '*y*' es un defensor de *'t'*, tal defensor es derrotado por '*x*' pero a su vez este es defendido por '*v*'. *'t'* es defendido por un argumento defendido, mientras que '*z'* es defendido por un argumento derrotado. Mientras que '*t*' es *exitosamente defendido,* '*z'* no lo es. De modo que una defensa exitosa será aquella que permite creer en el argumento originalmente planteado, como en el ejemplo 1.1. Por su parte, una defensa no exitosa será aquella que (aun verificando las exigencias de la definición 3.2) no puede garantizar la creencia razonable en un argumento determinado. En el último ejemplo, '*x*' "defiende" a '*z*' pero tal "defensa" no restaura el estado epistémico de '*z*'.

Esto sucede porque hay derrota de derrotadores, hay consistencia, pero tal derrota no neutraliza el efecto de la derrota sobre el argumento defendido.

Esto lleva a, por un lado, plantear si es preciso restringir la noción de defensa a casos exitosos, o por otro, evaluar la utilidad que tendría emplear una noción amplia de defensa que incluya los casos no exitosos, en fin de cuentas, la pregunta es: ¿Se debe o no exigirle a la defensa el resultado de la aceptabilidad de un argumento? Ahora bien, para zanjar definitivamente la cuestión un análisis detenido es preciso. Del mismo modo que la cuestión sobre la derrota indirecta, sirva esto al menos como para plantear la necesidad de una futura discusión de la cuestión.

Esta sección inició con una pregunta: ¿qué significa defender un argumento? o ¿Cuándo un argumento puede considerarse defensor (defendido) de (por) otro? En principio, es obvio que defender o ser defendido implica que un argumento derrote al derrotador de otro, pero se ha visto que esto no alcanza para capturar la idea informal de defensa. Nadie, en un debate, por ejemplo, estaría dispuesto a emplear un argumento que objeta una objeción al punto de vista sostenido, y que al mismo tiempo desacredite lo que se pretende defender. Además de desacreditar objeciones, un defensor debe ser consistente con lo que se pretende defender, sin embargo, se ha visto que estas dos condiciones no alcanzan para capturar adecuadamente la idea de defensa. Una tarea ha quedado pendiente para trabajos futuros en relación a las *derrotas indirectas* y a la *defensa exitosa*.

Relevancia filosófica de los Sistemas Argumentativos

Los sistemas argumentativos son modelos formales desarrollados en el ámbito de la Inteligencia Artificial, en particular bajo el paradigma conocido como representación del conocimiento y el razonamiento (KR&R). Estos sistemas tienen como objetivo la representación del conocimiento y del razonamiento con vistas a poder obtener inferencias plausibles a partir de información incompleta y potencialmente contradictoria. Claramente, y tal como ya lo afirmaba Loui (1987) en uno de los artículos fundacionales de los sistemas argumentativos, los aportes de la filosofía en el enfoque son de suma importancia. Ahora bien, de aquí no se sigue que sean relevantes para la filosofía. A continuación se intentará brindar algunas razones que puedan justificar la relevancia filosófica de los Sistemas Argumentativos.

Según Pollock, como lo ha remarcado en varios trabajos (1987, 1995), la filosofía puede verse beneficiada por la Inteligencia Artificial, y por los sistemas argumentativos en particular, porque se pueden hacer implementaciones computacionales de las teorías propuestas en los "*escritorios*", y en consecuencia, también ser puestas a prueba. La razón

de ello es que los sistemas argumentativos exigen asumir y formalizar una concepción del conocimiento y de los patrones de inferencia rebatibles razonables. Los Sistemas Argumentativos son modelos del razonamiento rebatible. Las discusiones al estilo de Toulmin (1958), Lehrer y Paxson (1969), Kyburg (1983), podrían verse expresadas bajo una *teoría formal de la argumentación rebatible.*

Adicionalmente, puede decirse que los sistemas basados en argumento dan herramientas para la consideración y discusión de un amplio abanico de temas relacionados con el conocimiento (Sudduth, 2011) la evidencia y la justificación (Kelly, 2014; Pollock, 1994), la teología (Plantinga, 2000) el razonamiento legal (Prakken, 1993), la negociación (Tomhé, 2002), como algunas contribuciones al terreno de la teoría de la argumentación (Marraud, 2007). Ya desde el plano de la lógica filosófica, los sistemas argumentativos, tal como lo señalan Prakken y Vreeswijk (2002) son formalismos capaces de atender a un amplio número de razonamientos ampliativos, tales como la inducción, analogía y la abducción. Además de esto han sido capaces de dar cuenta de la mayoría de los sistemas basados en lógica para razonamiento default o autoepistémico (Dung, 1995).

Anteriormente se ha destacado que los sistemas argumentativos ofician de modelo para la comprensión del razonamiento rebatible. Obviamente, que tal modelo supone o brinda una teoría de la retractabilidad y de la justificación de una creencia. Las nociones de derrota y defensa son claves en la comprensión de tal teoría. Ahora bien, será interesante decir algo al respecto de la distinción conflicto-derrota y de la importancia de pretender brindar una noción de defensa como la pretendida en este trabajo. Pero antes de ello cabe señalar que la afirmación filosófica más relevante de los sistemas argumentativos consiste en que para considerar si un argumento justifica la conclusión que sustenta, es necesario conocer únicamente las relaciones de derrota con otros argumentos (Dung, 1995). Ahora bien, conocer o aclarar la naturaleza de la derrota es clave para que no aparezcan problemas en el proceso de selección de los argumentos justificados.

Conocer la naturaleza de la derrota es importante puesto que como se ha señalado más arriba, la justificación de un argumento dependerá sólo de tal noción, de modo que si el concepto es inadecuado, los resultados se acarrearán al proceso justificatorio también. Por otro lado, es importante decir que la distinción conflicto-derrota es relevante porque no es lo mismo exigir a una teoría que sea libre de conflicto que de derrotas. Obviamente que una teoría libre de conflictos será libre de derrotas pero no viceversa.

Otro aspecto a considerar se puede ilustrar mediante una mirada práctica. Podría decirse que la naturaleza de ambas relaciones es distinta y puede

expresarse diciendo que el conflicto o el ataque es una *pretensión de derrota*. Un agente puedo tener la intención de socavar o refutar un argumento pero podría no tener éxito en la tarea, la derrota, en cambio, es *pretensión de derrota* y *éxito*. De lo contrario bastaría con construir cualquier argumento que interfiera con otros, i.e. *que esté en conflicto con aquellos*, para poder desacreditarlos. Es posible por cierto que haya casos de derrotas exitosas aparentes y derrotas exitosas genuinas, en el sentido de que puede ser que un argumento derrote a otro pero a su vez este se vea derrotado.

Con vistas a ilustrar la idea de pretensión de derrota y derrota, considere por caso, nuevamente, el ejemplo de Tweety. Supóngase que Juan sabe que Tweety es un pingüino y sabe que por lo general los pingüinos no vuelan, razón por la cual cree que Tweety no vuela, todo esto en el tiempo *t1*. Ahora bien, supóngase que se acerca Pedro y le dice: *así que ahí tienes un ave llamada Tweety, lo sé porque es un pingüino y todos los pingüinos son aves, y dado que es un ave, poseemos buenas razones para creer que Tweety vuela*, todo esto en el tiempo *t2*.

En esta situación, es claro que las conclusiones inferidas por Juan y Pedro son conflictivas, ya desde un punto de vista racional no es adecuado aceptarlas conjuntamente. Ahora bien, el ejemplo permite ilustrar que, conceptualmente, no es lo mismo conflicto que derrota, y la naturaleza de la derrota es diferente a la del conflicto, aunque en ocasiones esta diferencia sea por demás sutil. Es evidente que el argumento de Pedro no puede considerarse como un derrotador del de Juan a pesar de haberse esgrimido o formulado en un tiempo posterior, o haberse realizado con la *pretensión de derrotarlo*.

Derrotar a un argumento parece significar lo siguiente: Si de dos argumentos que no pueden ser aceptados mutuamente y uno de ellos es claramente mejor, se dirá que tal argumento es un derrotador del otro. Ahora bien, aunque esta noción parece clara, el ejemplo de Nixon propuesto más arriba, supone que o bien ambos argumentos están en conflicto y no hay derrota, o la derrota debe entenderse de una manera más amplia, por ello se propuso la definición 2.1.

La relevancia de la distinción conflicto-derrota puede también destacarse considerando el ejemplo 2.8 sobre el violín Stradivarius. Este ejemplo generaba resultados contraintuitivos en el modelo de Amgoud y Cayrol (1998) razón por la que debieron realizar algunas modificaciones en Amgoud y Vesic (1999). En este ejemplo existe una relación de ataque asimétrica (conflicto) pero no por ello debe considerarse como una derrota asimétrica. No parece tener sentido que el argumento sustentado en una frase de un infante constituya evidencia suficiente para negar el sustento

que llevan a asumir que el violín es caro. El argumento basado en lo dicho por el niño no derrota al argumento basado en la opinión de un experto en violines. Obviamente que queda un problema abierto aquí ¿quién derrota a quién? Si es que alguien derrota a alguien. Tal situación plantea la necesidad de realizar una clarificación del concepto que permita explicar estos casos.

Si la relación de derrota no es descompuesta en relaciones más elementales, no podría distinguirse las diferencias en los tipos de derrota. Esto es importante porque hay preferencias basadas en aspectos *internos* al argumento, como la especificidad, o por aspectos externos al argumento, *como pueden ser los valores o gustos de los agentes*. Podría parecer no muy grave, sin embargo, considere el ejemplo de Tweety nuevamente pero con el agregado de que Juan ha informado que Tweety es un Pingüino y que Pedro ha dicho que es un ave. Supóngase que Juan es un mentiroso compulsivo. Bajo esta nueva situación, aunque un argumento es más específico, el otro está basado en información más confiable por lo que la derrota no se da basada en especificidad sino basada en la confiabiliad de la fuente. En este caso la derrota no es del más específico al más general sino al revés.

Finalmente, vale referirse a un trabajo reciente, en el que Bodanza (2015) discute posibles consecuencias que trae aparejado diversas maneras de interpretar la noción de derrota. Puntualmente, en tal trabajo se señala que dependiendo de la concepción de derrota que se tenga es posible afirmar que los procesos justificatorios propuestos en (Dung, 1995); presentan problemas o si se asume que son razonables, las relaciones de derrota están mal concebidas.

Obviamente, puede advertirse que el concepto de defensa es tan importante como el de derrota. En el presente trabajo se ha pretendido brindar un acercamiento a la noción de defensa, esto se encuentra motivado por la pretensión de encontrar condiciones bajo las cuales una *creencia* (enunciado, argumento) epistémicamente desacreditada pase a considerarse razonable. Además, con el esclarecimiento de tal noción se pretende evitar la confusión que puede resultar de considerar a los *argumentos no derrotados* y los *argumentos no derrotados finalmente*. Estos últimos argumentos podrían llamarse más bien, argumentos defendidos exitosamente, puesto que en el fondo, p.e. en el caso de Tom Rabit, el primer argumento esta derrotado, pero sin embargo se puede creer en él, porque esta defendido exitosamente.

La realización de un estudio sobre los conceptos de derrota y defensa tienen en el fondo una intención filosófica bien clara: capturar o entender la *lógica* del *"a pesar de",* esto es de la defensa, y la lógica del *"a menos que",* esto es de la derrota. La derrota establece que un argumento deja de

ser razonable cuando se ha actualizado la, de algún modo tácita advertencia: esto es razonable *a menos que* tal y tal cosa. La defensa por su parte, es la actualización de la *desactivación de la advertencia* señalada anteriormente, la defensa establece que *a pesar de tal y tal cosa* (las objeciones o derrotas) tal argumento es razonable.

Conclusión

En el presente trabajo se ha pretendido realizar un acercamiento a las relaciones que pueden dares entre argumentos rebatibles: *derrota* y *defensa*. Se han discutido los conceptos en base a ejemplos y se han propuesto nociones formales de ambos conceptos. Específicamente dos cuestionamientos motivaron el presente trabajo: qué significa que un argumento derrota a otro y qué significa que un argumento defienda a otro. La respuesta a ambas preguntas también llevo a ciertas aclaraciones en ambos conceptos que sirvieron para clarificarlos.

Como conclusión se ha dicho que, mientras que para que haya derrota es preciso que ambos argumentos no puedan ser aceptados conjuntamente (obviamente, que no solamente), para que haya defensa, ambos argumentos deben poder ser aceptados conjuntamente. Se han planteado una serie de problemas para futuras indagaciones, principalmente con respecto a la relación entre derrota indirecta y defensa, y sobre si a la defensa debe o no exigírsele como resultado la aceptabilidad de un argumento.

Referencias bibliográficas

Alessio, C. A. (2015). *Restablecimiento y especificidad en sistemas argumentativos*. Doctoral dissertation: Universidad Nacional del Sur.

Amgoud, L., & Cayrol, C. (1997). "Integrating preference orderings into argument-based reasoning". In *Qualitative and Quantitative Practical Reasoning* (pp. 159-170). Springer Berlin Heidelberg.

Amgoud, L., & Cayrol, C. (1998, July). "On the acceptability of arguments in preference-based argumentation". In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence* (pp. 1-7). Morgan Kaufmann Publishers Inc.

Amgoud, L., & Vesic, S. (2009, July). "Repairing Preference-Based Argumentation Frameworks". In *IJCAI* (pp. 665-670).

Benferhat, S. (2000). "Computing specificity in default reasoning". En Gabbay, D.M. & Smets, P. (Eds.) *Handbook of Defeasible Reasoning and Uncertainty Management Systems* (pp. 147-177). Springer Netherlands.

Bodanza, G. A. (2015). "La argumentación abstracta en Inteligencia Artificial: problemas de interpretación y adecuación de las semánticas para la toma de decisiones". *Theoria: an international journal for theory, history and foundations of science,* 30(3), 395-414.

Bodanza, G.A. & Alessio, C.A. (2010) "Sobre la aceptabilidad de argumentos en un marco argumentativo con especificidad". *Actas de la II Conferencia Internacional Lógica, Argumentación y Pensamiento Crítico* (pp. 74-81). CEAR. Santiago, Chile.

Bodanza, G.A. & Alessio, C.A. (2014). "Reinstatement and the Requirement of Maximal Specificity in Argument Systems". *Logic, Language, Information, and Computation*: 81-93.

Bodanza, G. A., Tohmé, F., & Simari, G. R. (2012). "Argumentation Games for Admissibility and Cogency Criteria". En Verheij, B., Szeider, S., Woltran, S. (Eds.) *Computational Models of Argument. Proceedings of COMMA 2012* (pp. 153–164). IOS Press.

Chesñevar, C. I., Maguitman, A. G., & Loui, R. P. (2000). "Logical models of argument". *ACM Computing Surveys (CSUR)*, *32*(4): 337-383.

Dung, P.M. (1995). "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n* person games". *Artificial intelligence*, 77(2): 321-357.

Fuentes Bravo, C., & Santibáñez Yãnez, C. (2014). "Toulmin: razonamiento, sentido común y derrotabilidad". *Kriterion: Revista de Filosofía*, *55*(130): 531-548.

García, A. J., & Simari, G. R. (2004). "Defeasible logic programming: An argumentative approach". *Journal of Theory and Practice of Logic Programming*, *4* (1-2): 95-138.

Horty, J. F. (1994). "Some Direct Theories of Nonmonotonic Inheritance". In: Gabbay, D., Hobber, C., Robinson, J. (eds.) *Handbook of Logic in Artificial*

Intelligence and Logic Programming. Nonmonotonic Reasoning and Uncertain Reasoning, vol. 3 (pp. 111–187). Oxford University Press.

Horty, J. F. (2001). "Argument construction and reinstatement in logics for defeasible reasoning". *Artificial Intelligence and Law*, *9*(1): 1-28.

Horty, J. F. (2012). *Reasons as Defaults*. Oxford University Press.

Horty, J. F., Thomason, R.H., & Touretzky, D.S. (1990). "A skeptical theory of inheritance in nonmonotonic semantic networks". *Artificial intelligence*, *42*(2): 311-348.

Kyburg, H. (1983) "The Reference Class", *Philosophy of Science,* 50.

Koons, R. (2014). "Defeasible Reasoning", En Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*.

Lehrer, K & Paxson, T. (1969). "Knowledge: Undefeated Justified True Belief". *Journal of Philosophy*, *66*: 225-37.

Loui, R.P. (1987). "Defeat among arguments: a system of defeasible inference". *Computational intelligence*, *3*(1): 100-106.

Modgil, S., & Prakken, H. (2013). "A general account of argumentation with preferences". *Artificial Intelligence*, *195*: 361-397.

Plantinga, A. (2000). *Warranted Christian Belief.* New York: Oxford University Press.Plantinga applies his externalist theory of warrant and proper function to questions regarding the positive epistemic status of Christian belief. In chapter 11 Plantinga provides a more developed account of his view of rationality defeaters earlier introduced in Plantinga 1993a.

Pollock, J. L. (1987). "Defeasible reasoning". *Cognitive science*, *11*(4): 481-518.

Pollock, J. L. (1994). "Justification and defeat". *Artificial Intelligence*, *67*(2), 377-407.

Pollock, J. L. (1995). *Cognitive carpentry: A blueprint for how to build a person*. MIT Press.

Pollock, J. L. (2001). "Defeasible reasoning with variable degrees of justification". *Artificial Intelligence*, *133*(1): 233-282.

Poole, D. (1985). "On the Comparison of Theories: Preferring the Most Specific Explanation". En Joshi, A.K. (Ed.): *Proceedings of the 9th International Joint Conference on Artificial Intelligence* (pp. 144-147). Los Angeles, CA. Morgan Kaufmann.

Prakken, H. (1993). *Logical Tools for Modelling Legal Arguments*. PhD thesis, Vrije University, Amsterdam (Holanda).

Prakken, H. (1997). *Logical Tools for Modelling Legal Argument. A Study of Defeasible Reasoning in Law.* Dordrecht etc.: Kluwer Law and Philosophy Library.

Prakken, H. (2002). "Intuitions and the Modelling of Defeasible Reasoning: Some Case Studies". En Benferhat, S., Giunchiglia, E. (eds.) *9th International Workshop on Non-Monotonic Reasoning* (pp. 91-102). Tolouse, France.

Prakken, H. (2011). "An overview of formal models of argumentation and their application in philosophy". *Studies in Logic*, 4(1): 65-86.

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

Prakken, H., & Sartor, G. (1996[b]). "A system for defeasible argumentation, with defeasible priorities". En Gabbay, D.M. & Ohlbach, H.J. (Eds.) *Practical Reasoning, International Conference on Formal and Applied Practical Reasoning* (pp. 510-524). Bonn, Germany: Springer Berlin Heidelberg.

Prakken, H., & Sartor, G. (1997). "Argument-based extended logic programming with defeasible priorities". *Journal of applied non-classical logics*, *7*(1-2): 25-75.

Prakken, H., & Vreeswijk, G.A.W. (2000). "Logics for defeasible argumentation". En Gabbay, D. M. & Franz, G. (Eds.) *Handbook of philosophical logic,* Vol. IV (pp. 219-318). Springer Netherlands.

Simari, G.R. & Loui, R.P. (1992). "A mathematical treatment of defeasible reasoning and its implementation". *Artificial intelligence*, *53*(2): 125-157.

Sudduth, M. (2008) "Defeaters in Epistemology". In Fieser, J. and Dowden, B. (Eds.) *Internet* Encyclopedia of Philosophy

Tohmé, F. (2002). "Negotiation and Defeasible Decision Making". *Theory and* Decision, 53(4), 289-311.

Vreeswijk, G.A.W. (1993). *Studies in Defeasible Argumentation.* Doctoral dissertation: Free University Amsterdam.

Vreeswijk, G.A.W. (1997). "Abstract argumentation systems". *Journal of Artificial Intelligence*, *90*: 225-279.

Vreeswijk, G. A. W. & Prakken, H. (2000). "Credulous and skeptical argument games for preferred semantics". En Ojeda-Aciego, M., de Guzmán, I.P., Brewka, G. & Moniz Pereira, L. (Eds.) *Logics in Artificial Intelligence, European Workshop, JELIA Proceedings* , (pp. 239–253). Springer.

Walton, D. (2011). "How to refute an argument using artificial intelligence". In: M. Koszowy (ed.), *Argument and computation, Studies in Logic, Grammar and Rhetoric, 23*(36): 123–154.

Weisberg, J. (2015), "Formal Epistemology", En Edward N. Zalta (ed.) *Stanford Encyclopedia of Philosophy*.