¿Ropas nuevas para el Emperador?

Gladys Palau *

En su artículo *The Emperor's New Clothes* (1993), Sven Ove Hansson sostiene que los tres enfoques más importantes sobre condicionales contrafácticos, a saber, el correspondiente a la teoría de la derivabilidad (Chisholm, Goodman y Mackie), el de mundos posibles (Stalnaker y D.L. Lewis) y el de la teoría de revisión de Creencias (Harper, Levi y Gärdenfors), presentan los mismos problemas o dificultades recurrentes. En el presente trabajo sólo nos ocuparemos de los dos primeros enfoques analizados por Hansson en referencia a la dificultad de *no deducibilidad,* (i.e., utilizar elementos no deductivos) con el propósito de mostrar que las estrategias utilizadas en los sistemas lógicos analizados por Hansson se repiten aún en los formalismos no monótonos.

Dado que los tres enfoques analizados se inspiran en el *Test de Ramsey* (TR), consideramos necesario reproducirlo. Este dice:

(TR) ... *En general podemos afirmar junto con Mill que, (la afirmación) "Si p entonces q" significa "q es inferible a partir de p", esto es, por supuesto, a partir de p junto con ciertos hechos y leyes no afirmados pero de cierta forma indicados por el contexto. Esto significa que* $p \rightarrow q$ *se sigue de estos hechos y leyes.* (1931, p.248).

Nótese que RT, en tanto instrucción para evaluar un enunciado condicional propone, por un lado, que el enunciado *p entonces q* debe ser evaluado en tanto inferencia de q a partir de p, y por el otro, que q es inferible a partir de p en conjunción con un conjunto de enunciados implícitos pero sugeridos por el contexto. Como habrá notado el lector, la frase *p en conjunción con un conjunto de enunciados implícitos pero sugeridos por el contexto,* es la que evidencia el problema central de este tipo de lógicas, o sea, cómo incorporar en el formalismo este conjunto de enunciados de manera tal que en conjunción con el antecedente, conviertan a éste en una condición suficiente para el consecuente.

La primer propuesta para dar cuenta de las condiciones de verdad de un condicional contrafáctico basada en RT es la llamada teoría de la derivabilidad. Según la versión más acabada (Goodman 1947), las condiciones de verdad para los condicionales contrafácticos se establecen reduciendo las condiciones de verdad al concepto de deducibilidad clásico, de la siguiente forma:

(G) *Si A entonces B es verdadero si y sólo si B es derivable a partir de A junto con el conjunto de todas las leyes físicas y todas las proposiciones verdaderas cosostenibles (o lógicamente compatibles) con A.*

Expresadas en otros términos, la oración Si A entonces B es verdadera si y sólo si B es derivable a partir de A junto con el conjunto de todas las proposiciones verdaderas tales que ninguna de ellas implique contrafácticamente la negación

* Universidad de Buenos Aires.

Epistemología e Historia de la Ciencia, Volumen 10 (2004), Nº 10

de A y que la negación de algún miembro de ese conjunto sea contrafácticamente implicado por A.

Formalmente:

(G′) *A entonces B es verdadero si y solo si T/A ∪ {A} implica B, donde T/A es e un subconjunto de todas las proposiciones verdaderas que no implican ¬A (i.e., cosostenibles con A).*

Obviamente ni G ni G′ están formuladas en términos exclusivamente deductivos, ya que ellas implicas mecanismos de selección para el conjunto T/A que no lo son. En otras palabras, la petición formal de que el conjunto T/A no impliquen ¬A no alcanza para que el antecedente sea condición suficiente para determinar la verdad de un condicional contrafáctico, pues hay que determinar cuáles son los enunciados que no implican ¬A.

Hansson muestra también que la versión epistémica de la teoría de la derivabilidad formulada por Mackie (1965) tampoco evita este problema. En efecto, dada la siguiente formulación de Mackie en términos de conjuntos de creencias:

(M) *Hay un subconjunto K/A del conjunto previo de creencias K, tal que K/A no implica ¬A, y toda oración C de la forma si A entonces B es sostenible si y sólo si K/A∪{A} implica C.*

es fácil observar que ella reproduce la misma dificultad si se reemplaza en (G′) el conjunto T por el conjunto de creencias K y el subconjunto T/A por el correspondiente subconjunto de creencias K/A.

Hansson pasa luego a mostrar que en el enfoque basado en la semánticas de mundos posibles, ya sea bajo análisis de R. Stalnaker (1968) o D. Lewis (1973), se presenta el mismo problema Como es sabido, la propuesta de Stalnaker introduce en el lenguaje de su sistema el signo > para los enunciados condicionales de la forma *Si A entonces B* y propone una interpretación ontológica del TR en términos de mundos posibles. En particular, la no deducibilidad en la propuesta de Stalnaker se manifiesta en su semántica a partir de la introducción de una función de selección que elige para validar el condicional el mundo más parecido al actual, tal como se manifiesta en las condiciones de verdad dadas para la valuación de enunciados condicionales:

(S) *Considérese un mundo posible en el que A es verdadero y en el que los otros aspectos difieren mínimamente del mundo actual. Luego, "Si A, entonces B" será verdadero sólo si B es verdadero en ese mundo posible.*

Más precisamente, a los efectos de determinar el valor de verdad de un enunciado condicional, incorpora al sistema una función-selección diádica *f*, que toma como argumentos una proposición y un mundo y como valor el mundo posible más similar al mundo en el que el enunciado es verdadero. A los efectos de que el mundo seleccionado no pueda ser cualquier mundo y que la función de selección capture la idea de *mínimamente diferente* o *máximamente similar* entre mundos, Stalnaker formula determinados requisitos que necesariamente involucran elementos no deductivos.

A los efectos de observar la similitud entre las estrategias de Goodman y Stalnaker (o si se quiere Goodman-Mackie-Stalnaker, Hansson muestra que para toda oración A, el conjunto $T/A \cup \{A\}$ de la teoría de la derivabilidad (G') coincide con el conjunto de mundos $f(A,mi)$ de la semántica de Stalnaker, o sea $f(A,mi) \leftrightarrow T/A \cup \{A\}$.

Por su parte, D. Lewis introduce en el lenguaje objeto el signo $\Box\!\!\rightarrow$ como primitivo para los condicionales contrafácticos. Su semántica se diferencia de la de Stalnaker en el hecho de que en ella los A-mundos a tener en cuenta para validar un contrafáctico de la forma $A \Box\!\!\rightarrow B$ no es uno solo, sino que los A-mundos constituyen una esfera de mundos, entre otras posibles. Así, su semántica se presenta como un sistema de esferas de mundos posibles ordenadas de acuerdo a un orden de *similaridad comparativa*, relación cuyo especificación es obviamente no deductiva.

Deseamos observar, además, que el recurso de introducir una función de selección en el lenguaje objeto de un sistema de lógica condicional no se agota en los sistemas analizados por Hansson, sino que, por el contrario, se reitera en otros, como por ejemplo, en la propuesta de Lennart Åqvist (1973). Para su construcción, Åqvist parte de una lógica modal con el operador necesidad como primitivo, y luego agrega el operador lógico "*", equiparable en su semántica a la función selección f de Stalnaker. Dada una oración A del lenguaje, *A podrá, tentativamente, leerse como "A, bajo tales circunstancias", o "A, permaneciendo las otras cosas igual"o "A relevantemente con", de forma tal que el enunciado $A>B$ es definido como $*A \Rightarrow B$. Así, un enunciado condicional relevantes impliquen estrictamente. O sea:

(A) *$A>B$ será verdadero si y sólo si $*A \Rightarrow B$ es verdadero.*

Puede constatarse fácilmente que el conjunto T/A de (G') (o K/A de (M)) coincide con el conjunto *A de (A) respecto de la necesidad de completar el antecedente del condicional y en las dificultades que presenta en relación con la determinación de sus componentes no deductivos.

Posteriormente, en la década de los '80, los llamados condicionales derrotables emergieron como el nuevo emperador en el campo de las investigaciones sobre lógica condicional. Dada la estrecha relación existente entre los condicionales contrafácticos y derrotables en cuanto a su caracterización lógica, era de esperar que los nuevos sistemas condicionales proporcionaran nuevas estrategias que permitieran superar la objeción de no deducibilidad señalada por Hansson. Veremos de ahora en más cuán lejos se está de ello.

Los sistemas sobre condicionales derrotables abundan hoy en la literatura sobre el tema. En particular, nosotros nos referiremos sólo a las propuestas sobre los condicionales derrotables de James Delgrande (1978), Carlos. E. Alchourrón (1996) y D. Makinson, (1993,1994).

Previamente a la presentación de estas teorías, deseamos recordar las semejanzas y diferencias esenciales entre condicionales contrafácticos y condicionales derrotables. Recuérdese, además, que un condicional es derrotable cuando el agregado de nueva información en el antecedente puede "derrotar", i.e., hacer falso, al consecuente. Tanto los condicionales contrafácticos como los derrotables no validan las leyes lógicas conocidas como Refuerzo del Antecedente (RA), i.e.,

$(A\rightarrow B) \rightarrow ((A\wedge C)\rightarrow B)$, Contraposición, i.e., $(A\rightarrow\neg B) \rightarrow (B\rightarrow\neg A)$ y Transitividad, i.e. $((A\rightarrow B) \wedge (B\rightarrow C)) \rightarrow (A\rightarrow C)$. Sin embargo, a diferencia de los condicionales derrotables, los condicionales contrafácticos también validan el *Modus Ponens* Condicional (MPC): $(A\wedge(A>B)) \rightarrow B$ (o en su formulación alternativa: $(A>B)\rightarrow(A\rightarrow B)$).

El primero en construir una lógica para los condicionales derrotables fue James Delgrande, en su trabajo de 1987. Su idea era formalizar enunciados condicionales similares a *Todos los pájaros vuelan*. Evidentemente, éste es un condicional derrotable, ya que podría darse el caso de que apareciera un pájaro que no volara. Por ello, siguiendo la propuesta de Reiter, propone leerlo *En el curso normal de los eventos todos los pájaros vuelan* o, más brevemente, *Normalmente los pájaros vuelan* o *Típicamente los pájaros vuelan*. A este fin, introduce en el lenguaje objeto de la lógica de primer orden el signo ~> para representar precisamente al condicional derrotable.

Lo interesante de la presentación de Delgrande es que, a diferencia de los casos anteriores que destacan la falla de RA, él usa el ejemplo citado para demostrar que el primer principio que hay que eliminar es el *Modus Ponens*. En efecto, esta regla es la responsable de que, a partir de las proposiciones *Tweety es un pájaro* y *Todos los pájaros vuelan*, se siga *Tweety vuela* y de que, si se agrega la proposición *Tweety es un pingüino* se infiera la negación de la anterior conclusión, i.e., *Tweety no vuela*. En síntesis, un condicional derrotable tampoco debe satisfacer CMP. Sin embargo, esta diferencia sintáctica no evita que se repita la objeción de Hansson. En efecto, para dar las condiciones de verdad para un condicional derrotable, Delgrande construye una semántica en la cual los mundos posibles son estados y la relación de accesibilidad es una relación de normalidad entre estados, en el sentido de que un estado m_1 es menos excepcional que otro estado m_2 o que el estado m_1 es más normal que el estado m_2. Formalmente:

(D) A ~>B es verdadero en el estado m si y sólo si en los mundos menos excepcionales en los que A es verdadera, también es verdadera B

Si se comparan (S) y (D) se constata que la estrategia consiste nuevamente en adoptar el lenguaje y el poder deductivo de la lógica clásica y trasladar a la semántica el problema de determinar cuáles son los estados menos normales o más excepcionales para luego dar las condiciones de verdad para el condicional.

Pasaremos ahora al sistema DFT para los condicionales derrotables, tal como es presentado por C.Alchourrón en *Detachment and Defeasibility in Deontic Logic* (1996). El sistema DFT, inspirado según sus propias palabras en la ya mencionada propuesta de Åqvist,

Este sistema está construido sobre el sistema T de lógica proposicional modal más la conectiva para el condicional derrotable > y el símbolo de función *f*. Este operador es introducido a fin de representar en el sistema la idea intuitiva de que en el lenguaje común las oraciones de la forma A entonces B no se usan en el sentido de que el antecedente A es condición suficiente para B, sino que B se infiere a partir de A juntamente con un conjunto de supuestos determinados por el contexto de uso de un supuesto agente racional. Así, si $A_1,\ldots,A_n$ son las suposiciones asociadas con A, entonces *f*A debe leerse como la afirmación conjunta de A con todas las $A_1, \ldots A_n$ asociadas a A. Así, es posible establecer las condiciones de verdad para los condicionales derrotables de la siguiente manera:

(AL) A>B es verdadero si y sólo si $fA \Rightarrow B$ es verdadero

Las similitudes entre (G'), (G''), (M) y (A) son evidentes. En efecto, tanto el antecedente fA de la formulación de Alchourrón (AL) como el *A de la teoría de Åqvist (A), pueden entenderse como un subconjunto C/A del conjunto de todas las circunstancias que en unión con {A} implican al consecuente, tal como se afirma en las formulaciones (G'), (G'') y (M) de la teoría de la cosostenibilidad. Esta afirmación puede constatarse fácilmente si en (A) y (AL) se reemplaza fA y *A por C/A respectivamente. La diferencia entre la implicación material empleada en (G') y (G'') y la estricta utilizada de (A) y (AL) no es relevante respecto a nuestro propósito ya que ella refleja solamente una elección de la base lógica sobre la cual se construye el sistema. Finalmente, si en la semántica de Stalnaker se tomara el conjunto $f(A,mi)$ como el conjunto de circunstancias en los cuales los elementos del conjunto C/A son verdaderos, es posible mostrar también que dichos conjuntos coinciden. Más aún, desde la sintaxis, en todos estos sistemas el MPC, i e., $(A\wedge(A>B))\rightarrow B$ (o, en la formulación alternativa $(A>B)\rightarrow(A\rightarrow B)$), es una regla válida y todos satisfacen el Test de Ramsey.

Pasaremos ahora a analizar nuestro problema desde las formulaciones de los condicionales derrotables y contrafácticos que se han dado desde las teorías sobre consecuencia no monótona, creadas para dar cuenta de los procesos inferenciales de los razonamientos de sentido común. Se coincide que fue Dov M. Gabbay quien, en 1985, estableció por primera vez las propiedades generales del razonamiento no-monótono en su trabajo *Theoretical foundations for non-monotonic reasoning in expert systems*. Años después, en 1990, aparece el trabajo de obligada referencia en esta problemática, de S. Krauss. D. Lehmann y M. Magidor (KLM), titulado *Nonmonotonic Reasoning, Preferential Models and Cumulative Logics*, y, en 1994, la más acabada versión de los trabajos sobre este tema, *General Patterns in Nonmonotonic Reasoning* de David Makinson. Pese a que no hay una sola noción de consecuencia no monótona, todas ellas coinciden en introducir la relación de inferencia no-monótona o, si se desea, derrotable o plausible, con el signo $|\!\sim$ de forma tal que la expresión $A|\!\sim B$ debe leerse como: *B se sigue comúnmente de A*, o *B se infiere no monótonamente de A*, o *B se sigue plausiblemente* (o *derrotablemente*) *de A*. Al igual que las lógicas condicionales modales vistas, los sistemas se construyen a partir de la lógica clásica (i.e., son supraclásicos) añadiendo un conjunto de reglas de inferencia, entre las cuales las siguientes son las más importantes:

Reflexividad	*Corte o Transitividad Cumulativa* (TC)	*Monotonía cautelosa o cumulativa* (MC)
$A\mid\!\sim A$	$\dfrac{A\mid\!\sim B \quad A\wedge B\mid\!\sim C}{A\mid\!\sim C}$	$\dfrac{A\mid\!\sim B \quad A\mid\!\sim C}{A\wedge B\mid\!\sim C}$

Transitividad Cumulativa afirma que si de dos oraciones A, B, en conjunción, se infiere plausiblemente C y si, además, B se infiere plausiblemente de A, entonces C también se infiere plausiblemente de A. En otras palabras, si uno aumenta el conjunto de las premisas con una información que es una consecuencia de ellas, se siguen obteniendo las mismas conclusiones. Monotonía cautelosa se introduce en reemplazo de la propiedad de Monotonía, esencial a la deducción clásica.

En primer lugar, deseamos destacar que por introducir una consecuencia no-monótona, i.e., no deductiva, como condición necesaria para el análisis de los condicionales derrotables, hace que desde el principio se acepte como indispensable para el análisis de los mismos una formalización no deductiva. Pero, similarmente a los sistemas de Stalnaker, D. Lewis y Delgrande, la estrategia consiste en incorporar los criterios pragmáticos de selección a la semántica, tal como lo pasaremos a mostrar.

En líneas generales, las diferentes semánticas propuestas para estas nociones de consecuencia plausible tratan de rescatar la idea de McCarthy de dominios mínimos o extensiones mínimas de predicados. De ahí que, en lugar de todos los modelos, tal como se hace en los sistemas modales clásicos, se tome en cuenta solamente un tipo particular de modelos, según el tipo de *minimalidad* semántica que se elija. Por ello, estos modelos son llamados *modelos mínimos* o, si se tiene en cuenta el orden de preferencia entre estados, *modelos preferenciales* o, si se quiere, *modelos normales*, y de allí el nombre de *semánticas preferenciales*.

Esquemáticamente explicado, un modelo preferencial consta de un conjunto M de elementos arbitrarios, que pueden ser llamados *mundos* o *estados* (Makinson) o *estados epistémicos de un hablante* (KLM); una relación de satisfacibilidad $\models$ entre mundos y proposiciones del lenguaje que se comporta de acuerdo con las con las condiciones de verdad de las conectivas clásicas y, una relación $<_m$ de orden entre elementos de M, de forma tal que, $m<n$ debe leerse como n es más normal o típico que m, *o, el estado n es más preferido que el estado m*. Estas nociones permiten dar las condiciones de satisfacibilidad para una proposición de la forma A>B de la siguiente manera (Makinson, 1993):

> $m \models A>B$ sii $m' \models B$ para todo $m' \in M$ que sea *minimal* bajo la relación $<_m$ entre todos los mundos que satisfacen a A.

En palabras, A se satisface en un mundo si y sólo si B se satisface en todos los mundos que son minimales dentro del conjunto de los mundos que satisfacen A.

Pero, como también hay condicionales contrafácticos derrotables, Makinson propone las siguientes condiciones de verdad para ellos:

(MK) A>B es verdadero en el mundo m sii $m' \models B$, para todo $m' \in M$ que
a) es el más cercano a m entre los mundos más normales que satisfacen A, o
b) es el más normal respecto de m entre los más cercanos que satisfacen A.

A las similitudes antes señaladas respecto de las lógicas condicionales modales, debemos ahora añadir las correspondientes a las semánticas de las lógicas condicionales y las preferenciales. 1) En las contrafácticas, a fin de evaluar un condicional, se postula un orden entre mundos que permita elegir los mundos *mínimamente diferentes o máximamente similares* del mundo de referencia. Por su parte, teniendo en cuenta (D) y (MK) en las semánticas preferenciales o normales, los estados o mundos que se seleccionan son aquellos que se consideran *mínima-*

mente anormales o *atípicos o máximamente normales* respecto de un estado particular; 2) Según lo afirman (S), (D) y (MK), en ambos tipos de semánticas, se procede de un modo similar cuando se trata de validar un condicional. En las contrafácticas, a fin de evaluar el condicional A>B en un mundo *m* se seleccionan, entre los A-mundos, todos aquellos que difieran *mínimamente* de *m* que validen B; en las preferenciales, se toma, o bien el mundo más cercano a *m* entre los mundos más normales que satisfacen A, o bien el mundo más normal respecto de *m* entre los más cercanos que satisfacen A.

De lo expuesto, podría inferirse que la estrategia seguida en los sistemas no monótonos refleja en la semántica el mismo problema recurrente que las lógicas condicionales tratadas al comienzo, ya que en ella se reproduce el problema de la vaguedad propia de la relación *mínimamente atípico o anormal* entre estados o mundos. Sin embargo, a favor de esta última debe observarse que, desde su presentación, las dificultades señaladas han sido legitimadas como características propias de cualquier inferencia del sentido común. En síntesis, hemos visto que las soluciones ofrecidas a uno de los problemas recurrentes señalados por Hansson como insuperables se presentan bajo nuevas formas ahora legítimas. Por ello, podríamos concluir que el nuevo emperador se sigue vistiendo de ropas viejas, por más que sus formas parezcan novedosas.

Referencias

Alchourrón, C. (1996) Detachment and Defeasibility in Deontic Logic, en *Studia Logica*, 57, 5-18.

Delgrande, J. (1998) Conditional Logics for Defeseable Reasoning , en *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, vol. 3 Kluwer, 135-173.

Goodman, N. (1947), The Problem of Counterfactual Conditionals, en *Journal of Philosophy*, 44, pp. 113-128.

Hansson, S.O. (1993). The Emperor's New Clothes, Some recurring problems in the formal analysis of counterfactuals, en *Conditionals: from Philosophy to Computer Science*, Clarendon Press, 13-32.

Lewis, D. (1973). *Counterfactuals*, Blackwell, Oxford.

Mackie, J.L. (1965). Causes and Conditions, en *Causation and Conditionals*, Oxford University Press.

Makinson, D. (1993) Five Faces of Minimality, en *Studia Logica* 52, 339-379.

Makinson, D. (1994) General Patterns in Nonmonotonic Reasoning, en *Handbook of Logic in Artificial Intelligence and Logic Programming*, Clarendon Press.

Ramsey, F.P.,(1931). *The Foundations of Mathematics*, London Routledge & Kegan Paul.

Stalnaker, R. (1975). A Theory of Conditionals, en *Causation and Conditionals*, Oxford University Press.