CONDICIONALES CONTRAFÁCTICOS: CONDICIONES DE VERDAD Y SEMÁNTICA DE MUNDOS POSIBLES. ACERCA DE LAS TEORÍAS DE R. STALNAKER Y D. LEWIS*

GLADYS PALAU
Sociedad Argentina
de Análisis Filosófico

Introducción

Los problemas que plantea el análisis veritativo de las proposiciones condicionales tal como se dan en el lenguaje natural, llevan a pensar que el solo hecho de poseer antecedente falso no es condición suficiente para que la proposición condicional sea considerada verdadera. Ejemplos extraídos del uso común muestran que condicionales con antecedente falso son considerados a veces verdaderos y a veces falsos, según las creencias u opiniones que el hablante tenga respecto de sus contenidos significativos. Dificultades de este tipo han sido relevadas en una experiencia llevada a cabo por Matalón[1] —perteneciente a la escuela de Ginebra y colaborador de Piaget— que, entre otros resultados interesantes, muestra que la decisión frente a condicionales con antecedente falso, se dificulta tanto más, cuanto menor es la conexión significativa entre antecedente y consecuente. Frente a los condicionales "Si los elefantes son rosas entonces $2+2=4$" y "Si los elefantes son rosas, entonces $2+2=5$", los sujetos interrogados respondían en su mayoría que se trataba de proposiciones falsas o sin sentido, basándose fundamentalmente en la falsedad del antecedente. Aunque respecto del segundo caso, muchos sujetos se inclinaron a considerarlo verdadero, aludiendo que en un mundo tan absurdo como

* Trabajo realizado con la ayuda de una beca de la Sociedad Argentina de Análisis Filosófico.

[1] Matalón, B.: "Étude Génetique de L'implication" en Piaget y otros: *Implications, formalization et logique naturelle*, PUF, 1952.

aquél en que se afirma que los elefantes son rosas, todo podría resultar verdadero.

También es obvio que los llamados condicionales contrafácticos tampoco resisten el análisis en términos de la tabla de verdad del condicional material. Considerar verdaderos a todos los condicionales con antecedente falso, automáticamente vuelve verdaderos a los ejemplos más paradigmáticos de condicionales contrafácticos.

En los clásicos trabajos de C. I. Lewis, se estudiaba un condicional distinto del material, llamado implicación estricta. Pero tampoco estos estudios intentaban formalizar los condicionales de tipo contrafáctico u otros condicionales ordinarios pero con matices causales. Como veremos en este trabajo, varias de las leyes que valen tanto para el condicional material como para el estricto, resultan inválidas para muchos usos del condicional natural, sobre todo del que tiene matices contrafácticos. Por ejemplo, a partir de la afirmación condicional "Si raspo un fósforo, entonces se enciende" nadie deducirá la afirmación "Si raspo un fósforo y está humedo, se enciende", inferencia válida tanto para el condicional estricto como el material.

La situación así someramente descrita ha llevado a muchos lógicos a enfocar el problema del condicional natural y el contrafáctico desde una perspectiva no veritativa funcional, sin por ello renunciar a la posibilidad de definir las condiciones de verdad para una proposición condicional. La semántica moderna y especialmente la desarrollada a partir de Kripke para la interpretación de sistemas modales han permitido el desarrollo de nuevas teorías sobre el condicional. El análisis crítico de dos de ellas respecto de su adecuación al lenguaje natural constituye el propósito de este trabajo. Ellas son la teoría de Robert Stalnaker expuesta en "A Theory of Conditionals"[2] y la teoría de David Lewis tal como la desarrolla en su libro *Counterfactuals*.[3]

[2] Stalnaker, Robert: "A Theory of Conditionals", en *Causation and Conditionals*, ed. E. Sosa, Oxford University Press, 1975.
[3] Lewis, David: *Counterfactuals*, Oxford, 1973.

1. *La teoría de R. Stalnaker*

Su sistema toma en cuenta: (i) un conjunto K de mundos posibles; (ii) una relación de accesibilidad R entre los elementos de K, que es reflexiva pero que puede tener además otras propiedades que originarán distintos sistemas en forma similar a la lógica modal clásica; (iii) un mundo absurdo designado por "λ" en el cual todas las contradicciones son verdaderas y también sus consecuencias lógicas y el cual no es accesible para ningún mundo, ni de él se puede acceder a ningún otro; y (iv) una función-selección (binaria) *f* que tiene proposiciones y mundos posibles como argumentos y como valor un mundo posible. Esta función selecciona para cada antecedente A de una proposición condicional y cada mundo posible *i*, un mundo posible en el que A es verdadero. La condición de verdad es por lo tanto sencilla: una proposición condicional "A entonces B" es verdadera en el mundo *i* cuando el consecuente B es verdadero en el mundo seleccionado por la función de selección, o sea en f(A, i); y falsa cuando el consecuente es falso en ese mundo seleccionado. Representando por ">" la conectiva "condicional", tenemos:

A > B es V en *i*, si B es V en f(A, i)
A > B es F en *i*, si B es F en f(A, i)

Pero no cualquier mundo posible puede ser seleccionado por la función *f*, lo cual hace necesario establecer cuatro condiciones para ella. 1) Para todo antecedente A y mundo base *i*, *A* debe ser verdadera en f(A, i). Es decir, que para examinar el valor de verdad del condicional, es necesario analizar un mundo seleccionado en que el antecedente es verdadero, para luego ver qué pasa en ese mundo con el consecuente. 2) Para todo antecedente A y mundo base *i*, f(A, i) = λ si y sólo si no hay ningún mundo posible respecto de *i* en el que

[4] En adelante llamaremos "mundo antecedente" a todo mundo posible en el que el antecedente es verdadero. Similarmente, siendo A una proposición, llamaremos A-mundo a todo mundo posible en el que A sea verdadera.

A sea verdadero. Esto significa que si no hay ningún mundo en el que se cumpla el antecedente —como resulta en el caso de los antecedentes imposibles— entonces la función-selección elige el mundo absurdo. 3) Para todo mundo base *i* y todo antecedente A, si A es verdadero en *i*, entonces $f(A, i) = i$. Esto quiere decir que si la proposición condicional tiene un antecedente que es verdadero en el mundo base *i* (en cuyo caso la construcción contrafáctica ha sido usada quizás a causa de una creencia equivocada del hablante), la función-selección debe elegir ese mundo *i*. Puesto que, desde el punto de vista intuitivo, es legítimo pedir que el mundo seleccionado en el que se cumpla el antecedente sea el más similar o parecido al mundo actual, es natural que, para el caso de condicionales con antecedente verdadero, la función seleccione al mundo actual, puesto que no puede haber otro más parecido a él que él mismo. 4) Para todo mundo base *i* y todo antecedente B y B′, si B es verdadero en $f(B', i)$ y B′ es verdadero en $f(B, i)$ entonces $f(B, i) = f(B', i)$. O sea que si un antecedente B es verdadero en el mundo seleccionado donde B′ es verdadero y un antecedente B′ es verdadero en el mundo seleccionado donde B es verdadero, entonces el mundo seleccionado para B es el mismo que el mundo seleccionado para B′. De ahí que si una función-selección establece al mundo *j* como más parecido a *i* que el mundo *k*, ninguna otra función-selección podrá establecer a *k* como más parecido a *i* que *j*. Las condiciones 1 y 2 de *f*, sugieren que no hay otras diferencias entre el mundo real y el seleccionado fuera de las implicadas "implícitamente" o "explícitamente" por el antecedente y que la función-selección está basada en un orden de mundos posibles según su parecido al mundo real. Las condiciones 3 y 4 establecen precisamente un orden total entre los mundos posibles, en el cual el mundo real precede a todos los demás. Stalnaker considera que pueden agregarse condiciones pragmáticas que especifiquen la función-selección, pero las establecidas son suficientes para definir las nociones semánticas de validez y consecuencia para la lógica del condicional.

Stalnaker establece una serie de siete axiomas que ubican al operador condicional entre la implicación estricta y el condicional material, o sea que $\Box(A \supset B)$ implica $(A > B)$, que implica $(A \supset B)$. Los resultados de la teoría serán tratados en 3.

La idea central de la teoría es la siguiente: el conectivo condicional ">" así definido, posibilita hablar, desde el lenguaje objeto de la lógica modal, sobre situaciones posibles particulares no-actuales (no-reales). Las proposiciones condicionales son entonces afirmaciones sobre situaciones posibles que quizás no pertenezcan al mundo real. Para Stalnaker, esto mismo son los contrafácticos: afirmaciones sobre situaciones que no pertenecen al mundo real, es decir situaciones contrafácticas. Las proposiciones condicionales indicativas y las contrafácticas tienen por lo tanto una misma naturaleza: *ambas hablan de situaciones posibles no reales.* Veamos ahora cómo las condiciones de verdad propuestas se aplican similarmente a ambos tipos de condicionales. Sea el siguiente par de proposiciones condicionales:

(1) Si Rusia entra en el conflicto China-Vietnam, USA utilizará armas nucleares, y

(2) Si Rusia hubiera entrado en el conflicto China-Vietnam, USA habría utilizado armas nucleares.

Ambos condicionales serán verdaderos en *i*, si es verdad que USA utiliza armas nucleares en el mundo seleccionado en el cual es verdad que Rusia entre en el conflicto China-Vietnam. Para ambos casos el mundo seleccionado será el más similar a todos respecto de *i* y será *i* mismo si en él el antecedente resulta verdadero. Podríamos preguntarnos entonces si dentro de la teoría de Stalnaker es posible diferenciar un condicional indicativo de un condicional contrafáctico. Para diferenciarlos habría que demostrar que la función *f* selecciona mundos antecedentes distintos para cada uno de ellos. Por lo ya dicho, se ve que no es así. Para el caso 1), el antecedente es verdadero en el mundo real *i*, *f* selecciona a *i* por la condición 3. Respecto del caso 2) si el antecedente del contrafáctico resulta verdadero en *i* (lo cual no

está prohibido y constituye el caso para el cual la formulación contrafáctica es algo inadecuada), también por la misma condición 3, *f* selecciona al mundo real. De resultar el consecuente verdadero en *i*, tanto el caso indicativo como el contrafáctico resultan entonces verdaderos. Si por el contrario, tanto el caso 1) como el 2) tuvieran el antecedente falso (lo cual haría más adecuado el uso contrafáctico de 2)), la *f*-función seleccionaría, para ambos, el A-mundo posible más parecido de todos a *i*; y si en él, el consecuente resulta verdadero, ambos condicionales resultarían también verdaderos. Por lo tanto un contrafáctico es verdadero en *i* si su correspondiente condicional indicativo es verdadero en *i*; y si el contrafáctico tiene antecedente verdadero, se reduce al caso de un condicional indicativo con antecedente verdadero. Parecería ser entonces que el sentido contrafáctico de un condicional queda reducido a una cuestión de "implicatura" del lenguaje[5] o bien a la intención que tiene un hablante de hacer saber mediante la construcción contrafáctica que él piensa que el antecedente es falso. Tenemos por lo tanto un solo condicional cuya formulación indicativa o contrafáctica depende de la opinión que el hablante tenga respecto de la verdad del antecedente.

El tratamiento unificado de ambos tipos de condicionales resultaría una ventaja para la teoría, si para todo par de condicionales (el indicativo y la respectiva construcción contrafáctica) se cumpliera que el valor de verdad es el mismo. Hay pares de condicionales en los que tal propiedad parece cumplirse, por ejemplo: "Si abandona la bebida, conseguirá empleo" y "Si hubiera abandonado la bebida, habría conseguido empleo". Pero un hermoso ejemplo debido a Adams y que cita Lewis en el texto donde desarrolla la teoría que analizaremos después, muestra que hay casos en los que esta propiedad no se cumple:

[5] El término "implicatura" es usado aquí en el sentido de H. P. Grice en su artículo "Logic and Conversation".

(1) Si Oswald no mató a Kennedy, entonces alguien lo hizo.

y

(2) Si Oswald no hubiera matado a Kennedy, entonces alguien lo habría hecho.

Obviamente nadie osará dudar de la verdad de 1) pero sin embargo, para alguien que crea que efectivamente la muerte de Kennedy fue un hecho cometido por un "fanático solitario", 2) se vuelve falsa. Pero las condiciones de verdad para el condicional dadas por Stalnaker hacen que ambos condicionales adquieran el mismo valor veriativo, dado que la *f*-función selecciona al mismo mundo antecedente para ambos casos, y si el consecuente es verdadero en él, tanto 1) como 2) son verdaderos. Pero si el consecuente fuera falso, ambos condicionales resultarían a su vez falsos. La teoría falla entonces para estos casos. Podría argumentarse que el ejemplo de Adams es poco común y que la teoría de Stalnaker da cuenta de la mayoría de los pares condicionales. Pero el caso planteado por Adams resulta más vulgar de lo que a simple vista parece. Ejemplos similares son: "Si Martín no rompió el jarrón, entonces otro niño lo hizo" y "Si Martín no hubiera roto el jarrón, otro niño lo habría hecho"; "Si Cervantes no escribió el Quijote, entonces otro lo hizo" y "Si Cervantes no hubiera escrito el Quijote, entonces otro lo hubiera hecho". Pareciera que lo común de la formulación contrafáctica en estos ejemplos, radica en que agrega una nota de necesidad respecto del hecho expresado en el antecedente, modalidad ésta que la respectiva formulación indicativa no recoge. Esta característica es un elemento de juicio más a favor de la distinción entre condicionales indicativos y contrafácticos. El ejemplo de Adams u otros similares, constituyen por lo tanto un golpe central a la teoría y parecería, que si bien es cierto que tanto las proposiciones condicionales indicativas como las contrafáticas pueden aludir a situaciones no reales (al menos para la creencia del ha-

blante), una teoría adecuada sobre las proposiciones condicionales indicativas y contrafácticas debe conservar la distinción entre ambos tipos de condicionales.

Las condiciones de verdad establecidas por Stalnaker no permiten tampoco distinguir entre otros tipos de condicionales contrafácticos como los siguientes:

(3) Si Oswald no hubiera matado a Kennedy entonces (necesariamente) otro lo habría hecho.

y

(4) Si Oswald no hubiera matado a Kennedy entonces (tal vez) otro lo habría hecho.

Un hablante que no estuviera totalmente convencido de que la muerte de Kennedy fue producto de un complot, podría muy bien rechazar 3) como verdadera, pero sin embargo aceptar 4). Pero la teoría de Stalnaker tampoco parece estar en condiciones de diferenciar ambos tipos de contrafácticos y recoger el sentido de posibilidad expresado en 4). Esta distinción es recogida en la teoría de Lewis que trataremos más adelante y a grandes rasgos podría establecerse así: si tomamos los mundos más parecidos al real en los que se cumple el antecedente de 3) y 4), podría afirmarse que 3) es verdadero si en *todos* esos mundos se cumple que otro mató a Kennedy; y que 4) es verdadero si en *alguno* de esos mundos otro mató a Kennedy. El "todos" y el "algunos" constituyen la diferencia de matices entre la necesidad y la posibilidad expresadas en 3) y 4).

La misma razón que hace indistinguibles 3) y 4) en la teoría de Stalnaker, conduce a que en la teoría se acepte como verdad lógica el llamado Principio del Tercero Excluido Condicional, que consideramos más adelante y respecto del cual pueden abrigarse serias dudas.

Que el mundo seleccionado por la *f*-función sea uno solo, plantea de por sí una dificultad. Recordemos que ese mundo debe ser el más parecido al mundo real entre los que cumplen con el antecedente. Es así porque el mundo antecedente

seleccionado sólo difiere del real en lo que implica "implícitamente" o "explícitamente" el antecedente, y lo demás sigue igual. Esto plantea entonces el problema: ¿pueden variar dos mundos sólo en lo que "atañe" al antecedente y lo restante permanecer igual? Esto podría suceder sólo en el caso de que las proposiciones expresaran hechos independientes. Pero desde el pensamiento natural, ¿podrían concebirse dos mundos que difirieran entre sí nada más en el hecho de que en uno ocurriera un terremoto y en el otro no?; ¿cuáles son las cosas implicadas "implícitamente" por un terremoto? Parecería que el enfoque de Stalnaker se basa en una creencia ingenua de que dado un mundo *i* y un enunciado A falso en *i* existe un solo mundo *j*, que es el más parecido a *i* de todos los mundos que satisfagan a A. Puede argüirse que quizás no existe un mundo *j* único, sino más bien muchos mundos posibles en los que se cumple A, que se parecen a *i* en el grado más alto posible y que difieren entre sí en otros detalles. El solo hecho de que en *j* se cumpla A (falso en *i*), hace que el mundo *j* difiera de *i* en aspectos distintos de A y esta diferenciación puede darse en más de una manera. Tomemos el enunciado A "Los canguros no tienen cola" que obviamente es falso en el mundo *i*. En el mundo *j*, A sería verdadero, lo cual arroja una primer diferencia respecto de *i*. Pero si A es verdadera, entonces tampoco puede seguir valiendo que todos los marsupiales tengan cola y que todos los canguros sean marsupiales. Es obvio entonces que hay formas distintas en que un A-mundo puede variar respecto del real: puede ser que no se cumpla la relación marsupial-tener cola, o que falle la conexión entre canguro y marsupial. Llamando a estas formas distintas en que un mundo *j* puede variar respecto del mundo real *i*, "variantes" de *j*, y si al menos dos de esas variantes de *j*, son tan parecidas a *i* una como la otra, resultaría entonces dudosa la unicidad del mundo más parecido a *i* en que se cumple A.

Este supuesto de unicidad es el que subyace en la teoría de Stalnaker y el que genera las dificultades antes apun-

tadas. Consideremos ahora una teoría en la cual se abandona explícitamente este supuesto y que por lo tanto supera las dificultades mencionadas.

2. *La teoría de David Lewis*

Siguiendo el enfoque de Stalnaker, David Lewis propone una teoría del condicional contrafáctico, aceptando la existencia de mundos posibles y su comparación según su grado de similaridad o parecido. Su teoría no pretende ser una teoría general del condicional —como lo era para Stalnaker— y, por lo tanto, en ella no se consideran los condicionales indicativos en general, ni aquéllos en subjuntivo que carezcan de sentido contrafáctico, como los subjuntivos futuros "Si mi equipo ganara el partido próximo, sería campeón", los cuales según Lewis se asemejan más a condicionales indicativos. Respecto de los condicionales indicativos y contrafácticos, Lewis afirma expresamente que se trata de dos condicionales distintos y no de uno solo cuya diferenciación depende de la opinión que el hablante tenga respecto de la verdad del antecedente. Para mostrarlo cita el ejemplo de Adams antes mencionado.

Puesto que el modo subjuntivo no es en castellano, ni en inglés, índice unívoco de condicional contrafáctico, usaremos siempre para la construcción contrafáctica el pretérito pluscuamperfecto del subjuntivo, el cual sugiere sin dudas la presuposición de la falsedad del antecedente. La construcción "Si hubiera dejado la bebida, habría conseguido empleo", expresa indudablemente la opinión del hablante acerca de la falsedad del antecedente. Hay casos también en que otro tiempo del subjuntivo adquiere sentido contrafáctico, como, por ejemplo, "Si Kennedy viviera, todavía sería presidente de EEUU". Pero aquí el sentido contrafáctico no depende del modo verbal, sino del conocimiento extralingüístico de la muerte de Kennedy. El mismo tiempo verbal subjuntivo pierde ese sentido en otros contextos, por ejemplo, "Si viviera, lo encontraríamos agonizando". De ahí que para

una formulación no ambigua del condicional contrafáctico sea preferible usar el giro propuesto.

Lewis introduce dos operadores contrafácticos que intentan abarcar los dos sentidos de necesidad y posibilidad contrafáctica, comunes en el lenguaje natural. Para ello introduce dos operadores: el operador al que Lewis llama "would" contrafáctico, simbolizado por "$\Box\!\!\rightarrow$", que se debe leer "Si se hubiera dado ————, entonces se habría dado ————" y el operador "might" contrafáctico, simbolizado "$\Diamond\!\!\rightarrow$" y que leeremos; "Si se hubiera dado ————, entonces podría darse ————" o "Si se hubiera dado ————, entonces podría haberse dado ————". Estos dos operadores son interdefinibles:

$$A \Box\!\!\rightarrow B =_{df} -(A \Diamond\!\!\rightarrow -B)$$
$$A \Diamond\!\!\rightarrow B =_{df} -(A \Box\!\!\rightarrow -B)$$

A partir del conjunto de los mundos posibles, donde el mundo real es uno de sus elementos, establece una asignación tal, que a cada *i* le corresponde una esfera de accesibilidad *Si*, tal que los elementos de *Si* son los mundos accesibles a *i*, que no difieren de *i* más allá de cierto límite. Los mundos accesibles a *i*, pero de mayor similaridad integrarán esferas "interiores" a *Si*, y aquéllos de menor grado de similaridad formarán esferas que contendrán a *Si*. El mismo Lewis describe esta inclusión jerárquica de cada esfera como una especie de astronomía ptolemaica. Habrá tantas esferas interiores a *Si* como grados mayores de similaridad comparativa y tantas esferas exteriores como grados menores de similaridad comparativa respecto de *i*. Para cada mundo *i*, hay entonces un conjunto de esferas $i que contiene todas las esferas de mundos accesibles a *i*.

Las condiciones de verdad para el operador "would" son:

A $\Box\!\!\rightarrow$ B es verdadero en i (de acuerdo al sistema de esferas $i), si y sólo si

(1) Ningún A-mundo pertenece a una esfera S de $i, o

(2) Alguna esfera S en $i, contiene al menos un A-mundo y A ⊃ B vale en todo mundo perteneciente a S.

(1) expresa el caso de verdad vacua: o bien no hay ningún mundo en el que el antecedente sea verdadero, o bien sólo es verdadero en algún mundo de alguna esfera exterior a $i, inaccesible por lo tanto desde *i*. En ese caso, A no es "sostenible" en *i* y todo contrafáctico con antecedente insostenible o imposible es vacuamente verdadero en *i*; (2) presenta el caso no vacuo, que constituye obviamente el más interesante. Expresa la idea de que un contrafáctico es no vacuamente verdadero en *i*, si el condicional A ⊃ B vale en todos los A-mundos más similares a *i*. Por lo tanto en el caso no vacuo, hay al menos una esfera "A-permisiva",[6] y todo A-mundo perteneciente a ella, valida el condicional A ⊃ B.

Lewis llama *supuesto límite*, a la suposición de que para todo mundo *i* y antecedente A sostenible en *i*, esto es, verdadero en algún mundo de alguna esfera de $i, hay una esfera A-permisiva que es la más similar a *i*. Esta suposición permite postular para una proposición A sostenible en *i*, la existencia virtual de un conjunto de mundos posibles A-permisivos, tal que no haya otros mundos A-permisivos más parecidos al mundo *i*. Si contrariamente a lo establecido por este supuesto, el orden entre los mundos antecedentes fuera un orden denso, en el sentido de que dado un mundo antecedente cercano a *i*, hubiera siempre otro todavía más cercano, no habría A-mundos del máximo de similitud con *i*. Por otra parte, la aceptación del supuesto límite, permitiría una formulación más sencilla de las condiciones de verdad: un contrafáctico es verdadero en *i*, si el consecuente es verdadero en todo mundo antecedente más cercano a *i*. Lamentablemente para la teoría, no es segura la verdad del supuesto límite, razón por la cual Lewis no lo adopta. Menos todavía acepta Lewis el supuesto que en la sección anterior criticamos a Stalnaker. Este supuesto, según el cual para cada propo-

[6] Llamaremos A-permisiva a toda esfera en la que al menos hay un A-mundo.

sición A y mundo *i*, existe un mundo *j* que es el A-mundo más parecido a *i*, por lo tanto, puede ser considerado como una forma extrema del supuesto límite, aquélla en la que la esfera de máxima similitud contiene sólo un A-mundo.

De las condiciones de verdad dadas por Lewis para el "$\Box\!\!\rightarrow$", se desprende que todos los contrafácticos con antecedentes imposibles resultan vacuamente verdaderos y los contrafácticos con antecedentes verdaderos se reducen a condicionales materiales, puesto que se postula que el conjunto $\{i\}$, que tiene a *i* como único miembro pertenece a $i y ningún otro mundo es más similar a *i* que *i* mismo.

Pasemos ahora a considerar las bondades y posibles dificultades que esta teoría presenta.

(i) En primer lugar, queremos hacer notar que la teoría de Lewis no pretende ser una teoría del condicional en el sentido unificado que lo proponía Stalnaker, tal como ya lo dijimos al comienzo de este parágrafo. En segundo lugar, constituye un mérito indiscutible de la teoría, que ésta permita distinguir entre el contrafáctico "would" y el "might". Asimismo la interdefinición propuesta parece adecuada porque no es fácil encontrar un contraejemplo que demuestre su inadecuación respecto de su interpretación en el lenguaje natural.

(ii) Podría sospecharse que la teoría es demasiado permisiva respecto del operador contrafáctico "might", en el sentido de que las condiciones de verdad establecidas para el mismo dificultan falsear la mayoría de estos contrafácticos. Las condiciones de verdad de este contrafáctico se derivan de su definición en términos del operador "would" y de las condiciones de verdad de éste. Las condiciones de verdad derivadas son:

A $\Diamond\!\!\rightarrow$ B es verdadero en un mundo *i* (de acuerdo a $i) si y sólo si:

1) algún A-mundo pertenece a S en $i, y

2) toda esfera S en $i, que contiene al menos un A-mundo, contiene al menos un mundo en el que A $\wedge$ B vale.

Se ve claramente que estas condiciones de verdad permi-

ten considerar como verdaderos a la mayoría de los contrafácticos "might", a excepción de aquéllos cuya afirmación fuera la negación de un contrafáctico "would" verdadero. Pero estos casos no son los más interesantes en relación al lenguaje natural. Los contrafácticos más comunes en este contexto son del tipo "Si los chinos hubieran entrado en el conflicto de Vietnam, podría haber estallado una tercera guerra mundial" y estos contrafácticos, aún dentro del lenguaje natural es muy difícil considerarlos falsos. De esto resulta que lo que a primera vista pareciera una dificultad, se torna un punto a favor para la teoría respecto de su adecuación al lenguaje natural. Asimismo, las mismas o análogas dificultades para falsear un contrafáctico "might" se encuentran para validar el contrafáctico "would". Pero nuevamente, también esta tarea es dificultosa en los contextos naturales, a menos que se trate de contrafácticos "would" con fuerza de ley.

(iii) Pero también hay puntos dudosos respecto de la adecuación de la teoría. Dadas las condiciones de verdad de la misma, todos los contrafácticos con antecedentes imposibles resultan vacuamente verdaderos. Lewis sostiene además que existen argumentos intuitivos que avalan tal decisión. El primero, según el cual todo es verdadero en un mundo en el cual las contradicciones lógicas son posibles, parece razonable y coincide con los resultados de la investigación de Matalón citada precedentemente. Su segundo argumento se refiere a que en el método de demostración por reducción al absurdo, se utilizan contrafácticos con antecedentes que niegan precisamente verdades matemáticas o lógicas y que sin embargo tales contrafácticos se consideran verdaderos en tales contextos.

Por último, Lewis reconoce que entre los contrafácticos con antecedentes imposibles, hay algunos que tiene sentido afirmarlos y otros que no. Por ejemplo, tiene sentido afirmar que "si existiera el mayor número primo *p*, entonces $p! + 1$ sería también un número primo", mientras que carecería de sentido asentir que "si existiera el mayor número primo, en-

tonces los sólidos regulares serían seis". Pero estas diferencias no se deben según Lewis a las condiciones de verdad de ambos, sino a razones conversacionales.

Sobre la base de estos argumentos, nosotros podríamos preguntarnos qué es lo que ellos legítimamente prueban. A nuestro entender, sus argumentos prueban a lo sumo que *algunos* de los contrafácticos con antecedentes imposibles son verdaderos. En efecto, si por "imposible" se entiende "lógicamente imposible" y además se entiende la relación de deducibilidad a la manera clásica, se obtiene el resultado propuesto por Lewis.[7] Pero si se amplía la extensión de lo que se considera antecedente imposible más allá de lo lógicamente imposible, surgen dificultades. Y puesto que la teoría de Lewis pretende ser una teoría de los condicionales contrafácticos, no puede dejar de abarcar los casos de imposibilidad física, que se presentan fuera de toda duda, como los "más contrafácticos de todos", si se nos permite la expresión. Pasemos a considerar los dos siguientes contrafácticos:

(1) Si Urano y Neptuno no hubieran estado sujetos a la gravedad, Leverrier habría descubierto a Neptuno a partir de las irregularidades de la órbita de Urano.
(2) Si la órbita de Marte hubiera sido circular, Kepler no habría encontrado una diferencia entre sus cálculos y las observaciones.

Es obvio que tanto el ejemplo (1) como el (2) son contrafácticos con antecedentes físicamente imposibles, por cuanto ambos antecedentes niegan lo que una ley física afirma como caso. Sin embargo (1) es decididamente falso, porque se sabe que Neptuno fue descubierto a partir de irregularidades en la órbita de Urano que sólo se explicaban mediante la hipótesis de la existencia de un planeta exterior y la ley de gravedad. Por el contrario (2) es verdadero, puesto que Kepler encontró tal diferencia entre observación y cálculo

[7] Esto es así porque, según una de las leyes de la implicación estricta, de una contradicción lógica se puede deducir cualquier proposición.

sólo porque la órbita no era circular; pues si hubiera sido circular, tal diferencia no hubiera existido. Pero la teoría de Lewis hace a ambos contrafácticos verdaderos y, a nuestro entender, para estos casos no vale la argumentación de que uno podría ser asentido y el otro no por razones conversacionales. En la teoría de Lewis ambos ejemplos resultan verdaderos porque las condiciones de verdad establecen que "A $\Box\!\!\rightarrow$ B" es vacuamente verdadero en *i*, no sólo cuando A es lógicamente imposible, sino también cuando es lógicamente posible pero en un mundo tan "alejado" de *i* que "escapa" a la esfera de accesibilidad $i. Entonces es legítimo suponer que algunos antecedentes A sean insostenibles en *i* precisamente por ir en contra de las leyes físicas de *i* y tal es el caso de los contrafácticos con antecedentes físicamente imposibles como 1) y 2).

Pero nuestra respuesta negativa respecto de 1) se reitera respecto de otros ejemplos similares y podría generalizarse de la siguiente forma: cada vez que un hecho B se registra como verdadero en el mundo real *i*, por efectos del cumplimiento de una ley L, puede esperarse que cualquier hablante enterado de tal circunstancia, considere falso al condicional contrafáctico "no-L $\Box\!\!\rightarrow$ B", pese a que no-L sea insostenible en *i* y por lo tanto vacuamente verdadero para Lewis.

Puede observarse entonces que no sólo los argumentos de Lewis a favor de la verdad en el caso vacuo son insuficientes cuando el antecedente no es lógicamente imposible, sino que en tales casos pueden encontrarse ejemplos que van contra su conclusión.

Aun cuando Lewis no haya tenido en cuenta ejemplos del tipo de los mencionados, es consciente de que su respuesta al problema de los contrafácticos con antecedente imposible es dudosa y menciona otras posibilidades para establecer las condiciones de verdad del operador "would", de tal manera que no sea verdadero por mera vacuidad o imposibilidad del antecedente en *i*.

(iv) En su artículo "A Causal Theory of Counterfac-

tuals",[8] Frank Jackson afirma que las condiciones de verdad de los condicionales contrafácticos establecidas por Lewis no son adecuadas. Para ello muestra que dichas condiciones de verdad hacen verdaderos a condicionales contrafácticos que son obviamente falsos, tal como hemos nosotros tratado de demostrar para el primer caso del parágrafo anterior. A este fin cita un ejemplo extraído de la mecánica cuántica: dado un evento *a*, está causalmente determinado que puede ocurrir *c1* o *c2*, pero cuál de ellos pueda ocurrir es totalmente azaroso. *a* puede ser interpretado como una cierta perturbación dentro de un átomo y *c1* y *c2* como órbitas resultantes de los electrones. Se sitúa en un tiempo *t* anterior a que *a* ocurra y se pregunta qué hubiera ocurrido si *a* hubiera ocurrido. Se presentan entonces posibles contrafácticos:

(1) Si *a* hubiera ocurrido, entonces *c1* o *c2* habrían ocurrido.
(2) Si *a* hubiera ocurrido, entonces *c1* habría ocurrido.
(3) Si *a* hubiera ocurrido, entonces *c2* habría ocurrido.

En física cuántica (1) es verdadero, mientras que (2) y (3) son falsos porque no se puede decidir nunca si ocurrirá *c1* o *c2*. La probabilidad de que en el mundo real *c1* haya ocurrido exactamente el mismo número de veces que *c2* es muy pequeña. Podemos suponer, pues, que en el mundo real *c1* se ha dado con una frecuencia mayor que *c2*. En ese caso, los mundos antecedentes en los que ocurre *c1* después de *a*, son más similares al mundo real que aquéllos en los que *a* está seguido de *c2*. Las condiciones de verdad hacen entonces a (2) verdadera, mientras que está demostrado en física cuántica que es falsa. Razonando similarmente se demuestra que las condiciones de verdad hacen verdadera a (3), si se parte del supuesto de que en el mundo real la frecuencia de que *c2* haya seguido a *a* es mayor que la de *c1*.

La crítica de Jackson va aún más lejos. En un mundo no determinista en el que los hechos particulares fueran tal cual

[8] En *Australasian Journal of Philosophy*, vol. 55, No. 1, 1977.

son, pero donde no hubiera causas y efectos (lo que él llama mundo humeneano) y donde la conjunción de dos eventos fuera puramente casual, ningún contrafáctico de los que él denomina *contrafáctico estándar* resulta verdadero.[9] Además, los contrafácticos "might" respectivos resultan todos verdaderos. Por el contrario, para la teoría de Lewis esto no podría suceder, porque existiría la posibilidad de decidir qué mundos accesibles son más similares a este mundo humeneano y de eso dependería el valor de verdad del contrafáctico analizado.

Jackson toma un ejemplo clásico de la literatura sobre los contrafácticos para demostrar que esta situación puede reproducirse en otros contextos. Dado el condicional contrafáctico "Si Bizet y Verdi hubieran sido compatriotas, ambos habrían sido franceses" y el contrafáctico "Si Bizet y Verdi hubieran sido compatriotas, habrían sido italianos", no puede decirse cuál de ellos es el verdadero, a menos que la investigación histórica agregue datos relevantes nuevos. Sólo podrá ser verdadero el contrafáctico "Si Bizet y Verdi hubieran sido compatriotas, ambos habrían sido franceses o ambos habrían sido italianos". Sin embargo, si seguimos a Lewis, la semejanza de algunos mundos posibles pueden llevarnos a verificar uno de los dos primeros ejemplos.

(v) Finalmente, creemos importante destacar una autolimitación aparentemente innecesaria en la teoría de Lewis. Mientras la teoría de Stalnaker pretendía dar una formulación unificada del condicional, Lewis elimina desde el principio los condicionales indicativos y los subjuntivos futuros, pues parece que los considera esencialmente distintos respecto de su comportamiento veritativo. Si bien el ejemplo de Adams ofrece una base inobjetable para considerar distintos los condicionales indicativos de los subjuntivos, Lewis no explicita las razones que lo llevan a eliminar los subjuntivos futuros. En principio, no parece haber inconvenientes en ex-

[9] Contrafáctico estándar es aquel que cumple las siguientes condiciones: (i)-p; (ii)-$\Box((p \supset q)$ y (iii) p y q no son contrafácticos.

tender las condiciones de verdad de Lewis a estos condicionales. En efecto, si un condicional subjuntivo futuro tuviera antecedente verdadero, al igual que un contrafáctico con antecedente verdadero, se reduciría a un caso del condicional material. Por otra parte, si un condicional subjuntivo futuro tuviera antecedente falso en el mundo actual *i*, para determinar su valor de verdad sería razonable elegir, dentro de una esfera de accesibilidad *Si*, el mundo *j* más parecido a *i* en el que el antecedente sea verdadero y luego ver qué valor de verdad tiene en ese mundo el consecuente, en forma totalmente similar a un condicional contrafáctico. Por ejemplo, supongamos el siguiente condicional subjuntivo futuro: "Si el Cha de Irán viviere diez años más, podría ser curado del cáncer." Puesto que el antecedente de este condicional es casi con seguridad falso en el mundo real, al igual que para un condicional contrafáctico, habrá que determinar si en todo mundo antecedente más parecido a *i*, se cumple el consecuente. Estas consideraciones parecen mostrar que la teoría de Lewis es susceptible de una extensión a casos interesantes que él parece haber dejado de lado.

3. *Comparaciones y resultados comunes*

A pesar de que ambas teorías se basan en la noción de mundos posibles y en la relación de similaridad comparativa entre los mundos posibles, podemos establecer las siguientes diferencias, varias de ellas remarcadas por el mismo Lewis.

1) La teoría de Stalnaker depende de un supuesto más fuerte que la teoría de Lewis. El supuesto de Stalnaker afirma que existe un solo mundo antecedente que es *el* más próximo a *i*. Este supuesto es aún más fuerte que suponer dentro de la teoría de Lewis el supuesto límite. El mismo Lewis afirma que la teoría de Stalnaker es un caso particular de la suya: el caso en que para todo mundo *i* y antecedente A sostenible en *i*, hay una *Si* que contiene exactamente un A-mundo. Esto implica el supuesto límite, pero no viceversa.

2) Stalnaker introduce la noción de mundo absurdo; Le-

wis demuestra que es técnicamente innecesaria. En efecto, la noción de un mundo absurdo sirve en la teoría de Stalnaker para establecer la verdad vacua de los condicionales con antecedentes imposibles; en la teoría de Lewis también se determina la verdad vacua para los contrafácticos con antecedentes imposibles sin necesidad de formular las condiciones en términos de un mundo absurdo.

3) Asimismo Lewis cree innecesario introducir la relación de accesibilidad en forma independiente, tal como lo hace Stalnaker, pues esta relación se puede definir en términos de la *f*-función.

4) Como ya lo dijimos en el parágrafo 1, excepto en el caso vacuo, la teoría de Stalnaker no puede distinguir entre los operadores "might" y "would", recogiendo sólo el sentido de este último.

5) Stalnaker caracteriza el conectivo condicional ">" mediante un sistema formal que no es necesario detallar aquí y que sitúa al conectivo ">" como intermedio entre el condicional estricto y la implicación material. Los condicionales contrafácticos o indicativos con antecedente verdadero se reducen a condicionales materiales, pero los condicionales indicativos o contrafácticos con antecedente falso pueden adquirir valores de verdad distintos del condicional material correspondiente. En efecto, si el condicional tiene antecedente falso y en el mundo seleccionado donde el antecedente es verdadero, el consecuente también lo es, el condicional resulta verdadero en el mundo actual. Pero, por el contrario, si el antecedente es falso en el mundo actual y en el mundo antecedente seleccionado el consecuente es falso, el condicional es falso en el mundo actual, a diferencia del condicional material que es verdadero para esos casos. Resulta pues, que dentro de la teoría de Stalnaker un condicional cuyo consecuente es falso en el mundo antecedente seleccionado, es siempre falso, independientemente del valor del antecedente en el mundo actual. Asimismo, si el consecuente es verdadero en el mundo antecedente, el condicional siempre es

verdadero independientemente del valor de verdad del antecedente en el mundo actual.

En la teoría de Lewis, puesto que los contrafácticos con antecedente verdadero se reducen a condicionales materiales, el operador "would" implica al condicional material, sin que esto quiera decir que el operador contrafáctico pueda ser definido a partir del condicional material.

6) En la teoría de Stalnaker es válido el principio del tercero excluido contrafáctico: $(A \Box\!\!\rightarrow B) \vee (A \Box\!\!\rightarrow \sim B)$, lo cual constituye para Lewis el "principal vicio y la principal virtud" de la teoría. En el caso vacuo ambos disyuntos son verdaderos; en el caso no vacuo, puesto que el mundo antecedente seleccionado por la función-selección es uno solo, de acuerdo al principio del tercero excluido común, o bien vale B, o bien vale no-B en tal mundo, y en esos casos el contrafáctico es verdadero o falso respectivamente. Lewis reconoce que la aceptación de este principio es plausible, porque gracias a él se justifica que las expresiones 1) $\rightarrow (A \Box\!\!\rightarrow B)$ y 2) $A \Box\!\!\rightarrow \sim B$ sean indiferenciables en el lenguaje natural. Tanto en la teoría de Stalnaker como en la de Lewis, excepto en el caso vacuo (2) implica (1), y suponiendo el principio del tercero excluido contrafáctico, (1) implica (2). Pero si se acepta este principio y la indiferenciación de (1) y (2), se pierde dentro de la teoría de Lewis la distinción entre el contrafáctico "would" y el "might", como él mismo sostiene. En efecto, en el caso no vacuo, si $(A \Box\!\!\rightarrow B)$ es verdadero, entonces $\sim(A \Box\!\!\rightarrow \sim B)$, y entonces por definición $(A \Diamond\!\!\rightarrow B)$; y recíprocamente, si $(A \Diamond\!\!\rightarrow B)$ es verdadero, entonces $\sim(A \Box\!\!\rightarrow \sim B)$ es verdadero y entones, por el principio del tercero excluido contráctico, vale $A \Box\!\!\rightarrow B$. Por lo tanto, ambos contrafácticos son interdeducibles, pese a que las condiciones de verdad que para ellos fija la teoría son distintas.

En la teoría de Lewis este principio no es válido porque ambos disyuntos pueden ser falsos. Esto es así, porque en su teoría se permite que haya más de un A-mundo con igual

similaridad comparativa y, por lo tanto, en un A-mundo B puede ser falso y en otro A-mundo $\sim$ B puede serlo.

Analicemos ahora el críptico ejemplo de Adams ya citado para ver cómo se comporta respecto de este principio. Sea entonces el enunciado "Si Oswald no hubiera matado a Kennedy, otro lo habría hecho" o "Si Oswald no hubiera matado a Kennedy otro no lo habría hecho". Ya dijimos anteriormente que el primer miembro puede ser falso, luego, si el principio del tercero excluido contrafáctico valiera, debería ser verdadero el segundo disyunto. Sin embargo, es plausible que pueda ser falso, porque un hablante que opine que la muerte de Kennedy no fue producto de una conspiración y que por lo tanto sostenga que el primer disyunto es falso, puede sostener sin contradicción alguna que el segundo disyunto también es falso, pues nada impide que otra persona que no sea Oswald hubiera matado a Kennedy. Puede verse además que las expresiones (1) y (2) que Lewis cree indiferenciables en el lenguaje natural, no se comportan como tales respecto de este mismo ejemplo, sino que por el contrario, es posible que adquieran valores de verdad distintos. Puesto que el primer miembro del principio del tercero excluido es falso, (1) se hace verdadera, mientras que (2) sigue siendo falsa. Del análisis de este ejemplo y de otros similares que pueden encontrarse, surge que la teoría de Lewis tiene más bondades y más matices que los que él mismo ha establecido a este respecto.

Pasemos ahora a considerar los resultados comunes a ambas teorías.

1) *Refuerzo del antecedente*

En ambas teorías sale como resultado la invalidez de la inferencia conocida con el nombre de refuerzo del antecedente: $A \mathrel{\square\!\!\rightarrow} B$, luego $(A.C) \mathrel{\square\!\!\rightarrow} B$. A título de contraejemplo para la validez de esta inferencia, Lewis cita uno similar al siguiente:

(1) Si el despertador hubiera sonado a las seis de la ma-

ñana, habría alcanzado el tren de las siete.

(2) Si el despertador hubiera sonado a las seis de la mañana, pero el auto no hubiera arrancado, no habría alcanzado el tren de las siete.

Como puede verse, si se refuerza el antecedente de (1), se obtiene otro condicional contrafáctico, que puede ser también verdadero, pero que tiene como consecuente la proposición contraria, y en este caso no se verificará el condicional reforzado pero con el consecuente inalterado. En forma similar podría construirse otro condicional contrafáctico, reforzando todavía más el antecedente:

(3) Si el despertador hubiera sonado a las seis de la mañana y el auto no hubiera arrancado, pero mi vecino me hubiera llevado, entonces habría alcanzado el tren de las siete.

Se podrían obtener así cadenas de condicionales contrafácticos en los cuales cada uno agregue un refuerzo más al antecedente y el consecuente sea siempre el consecuente del anterior negado. O sea:

(1) $A \mathrel{\square\!\!\rightarrow} B$
(2) $(A.C) \mathrel{\square\!\!\rightarrow} \sim B$
(3) $(A.C.D) \mathrel{\square\!\!\rightarrow} B$
etcétera.

Por su parte en la teoría de Stalnaker esta inferencia resulta también inválida, con el agregado de que también lo es para los condicionales indicativos.

Para desgracia del condicional material, este resultado parece adecuarse a la mayoría de los condicionales usados en el lenguaje natural, ya sean condicionales contrafácticos o indicativos. En efecto, para un hablante cualquiera, los tres ejemplos citados son verdaderos y, para determinar su verdad, se tomarán en cuenta mundos antecedentes distintos,

cada uno de los cuales será el mundo más parecido posible al real y en los que se cumpla el consecuente.

2) *Transitividad*

Lo grave del resultado anterior es que parece arrastrar la invalidez de la transitividad, que constituye tal vez una de las inferencias más intuitivas del pensamiento natural. Puesto que el refuerzo del antecedente se sigue de la transitividad, la invalidez del primero arrastra la invalidez de la segunda.[10]

Tanto en la teoría de Stalnaker como en la de Lewis, la invalidez de la transitividad surge además de las condiciones de verdad de las teorías. En la de Stalnaker, la función-selección puede seleccionar mundos antecedentes distintos para cada una de las premisas y la conclusión; en la de Lewis las esferas tomadas en cuenta pueden diferir de premisa a premisa, permitiendo la posibilidad de premisas verdaderas y conclusión falsa. Como contraejemplo para la validez de la transitividad contrafáctica, Lewis cita el siguiente ejemplo:

Si Otto hubiera venido, habría venido Ana.
Si hubiera venido Ana, habría venido Waldo.
Luego, si hubiera venido Otto, habría venido Waldo.

Bajo las suposiciones: 1) Ana va a todos los lugares donde Otto va; 2) Otto es el rival exitoso de Waldo respecto de los sentimientos de Ana; 3) Waldo sigue a Ana a los lugares donde ésta va, pero teniendo siempre cuidado de no encontrarse con Otto, y 4) Otto difícilmente podría concurrir a la fiesta por encontrarse fuera de la ciudad, las premisas se hacen verdaderas y la conclusión falsa. Pasemos ahora a detallar cómo, según Lewis, se verifican las premisas y se falsifica la conclusión. La primera premisa es verdadera porque en el mundo antecedente más parecido al real y en el

[10] Puede demostrarse fácilmente que el refuerzo del antecedente se sigue de la transitividad: supongamos $A \Box\!\!\rightarrow B$ y sea $A.C \Box\!\!\rightarrow A$ (caso de ley lógica); de ambas, por transitividad, se deduce $A.C \Box\!\!\rightarrow B$. Pero como se demuestra la invalidez del refuerzo del antecedente, luego también se invalida la transitividad.

que Otto haya ido a la fiesta, por el supuesto 1) Ana también va. La segunda premisa también es verdadera, porque en el mundo antecedente en el que Ana haya concurrido, Otto no habría ido (por 4) y Waldo sí (por el supuesto 3). Ambas premisas se verifican, pero debe observarse que se alude a mundos antecedentes distintos, puesto que en el primero Otto ha ido a la fiesta y en el segundo no. Finalmente la conclusión se falsifica, porque en el mundo antecedente más parecido al actual en el que Otto hubiera ido a la fiesta, por el supuesto 3) Waldo no habría ido. Stalnaker argumenta en forma análoga para demostrar la invalidez de la transitividad condicional y precisamente por este hecho hemos elegido para nuestro análisis el lenguaje y los supuestos de Stalnaker, que resultan más sencillos a fines expositivos.

Pareciera que la argumentación respecto del ejemplo de Otto, termina definitivamente con la transitividad contrafáctica. Pero, en primer lugar, queremos hacer notar que una argumentación similar termina también con la transitividad indicativa. Está claro que un resultado de este tipo no puede adjudicársele a la teoría de Lewis, por cuanto ella no pretende abarcar los condicionales indicativos. Pero este no es el caso de la teoría de Stalnaker, en la cual, como ya vimos, los condicionales indicativos aluden también a situaciones no reales y están sujetos a las mismas condiciones de verdad. El ejemplo anterior formulado en modo indicativo y las condiciones establecidas para la función-selección hacen evidente nuestra afirmación. En segundo lugar, si se examina el uso de los condicionales contrafácticos en el lenguaje natural, puede observarse que ellos se comportan respetando la transitividad. Pasaremos ahora a mostrar algunos ejemplos de inferencias contrafácticas que involucren transitividad y que difícilmente podrían demostrarse inválidas.

Citemos primero un ejemplo extraído de la experiencia que llevó a cabo Michelson para determinar la existencia del éter. Después de realizado el experimento y observar que la luz no había experimentado ningún retraso en su trayectoria, Michelson pudo haber razonado de la siguiente forma:

si el éter hubiera existido, entonces habría ofrecido un obstáculo al pasaje de la luz, y si hubiera ofrecido un obstáculo, entonces la luz habría experimentado un retraso en su trayectoria. Luego, si el éter hubiera existido, la luz habría experimentado un retraso en su trayectoria. Se sabe que como tal retraso no existió, Michelson concluyó la inexistencia del éter. (Otro ejemplo: supongamos la experiencia llevada a cabo por Gauss cuando midió los ángulos del triángulo formado por los rayos de luz enviados desde los picos de tres montañanas, a fin de probar la estructura del espacio físico; y supongamos un historiador de la ciencia reflexionando sobre tal experiencia: si la suma de los ángulos internos de un triángulo le hubieran dado a Gauss una diferencia muy grande, superior a los 180°, seguramente Gauss hubiera concluido que la geometría del espacio físico no era euclideana sino rimaneana, y en ese caso la curvatura del espacio habría sido mayor que cero. Luego, si la suma de los ángulos internos le hubieran dado una diferencia muy grande, superior a los 180°, Gauss hubiera considerado la curvatura del espacio mayor que cero.) Ejemplos similares pueden encontrarse sin dificultad en los análisis históricos, sociológicos, económicos, etc., cuando se quiere evaluar qué hubiera ocurrido o qué no hubiera ocurrido de haberse producido o de no haberse producido un determinado evento. Esta transitividad contrafáctica la podemos encontrar en casos mucho más cercanos a la vida cotidiana; ninguna persona que se haya interesado en el campeonato mundial de futbol jugado en la Argentina, encontraría incorrecta la siguiente inferencia: si Holanda hubiera logrado el gol en el minuto 43 del segundo tiempo, habría ganado el partido, y si hubiera ganado el partido, Argentina no habría salido campeón mundial; luego, si Holanda hubiera logrado ese gol, Argentina no habría salido campeón.

Si se analizaran estos ejemplos bajo la luz de las condiciones de verdad de las teorías de Lewis y Stalnaker, resultarían ser inferencias inválidas. Esto hace pensar que ambas teorías se alejan del comportamiento lógico de los condicionales con-

trafácticos habituales. La razón de ello es que las condiciones de verdad de dichas teorías se aplican a cada contrafáctico en forma independiente del otro, lo cual permite elegir los mundos antecedentes de cada contrafáctico en forma totalmente independiente uno del otro. Esto lleva directamente a "consultar" en una misma inferencia contrafáctica mundos que en algunos casos pueden ser totalmente opuestos, como en el caso de Otto, en el que en el primer mundo antecedente Otto ha concurrido a la fiesta y en el otro no.

Por el contrario, nosotros sostenemos que las condiciones de verdad de un enunciado contrafáctico no son independientes de ningún otro contrafáctico que se enuncie en el mismo contexto, en el sentido que ahora pasaremos a examinar.

Comenzaremos por analizar este requerimiento desde las exigencias del lenguaje natural. Todas las inferencias por transitividad contrafáctica tienen la siguiente forma: si $A \mathrel{\square\!\!\rightarrow} B$ y $B \mathrel{\square\!\!\rightarrow} C$, luego $A \mathrel{\square\!\!\rightarrow} C$. Todo lo que un hablante pide cuando enuncia una inferencia de este tipo es que una vez seleccionado el mundo antecedente A más próximo a *i* en el que se verifica B, o sea el mundo en el que la primera premisa es verdadera, en el mundo antecedente seleccionado para verificar si se cumple la segunda premisa, siga siendo verdadero A. Más brevemente, se pide que el B-mundo antecedente de la segunda premisa sea asimismo un A-mundo. Desde un enfoque distinto, podría afirmarse que el antecedente A del primer contrafáctico fija o retrotrae toda la situación contrafáctica a los mundos en los que él se verifica. Ejemplificaremos con una de las inferencias del lenguaje natural dada anteriormente. Sea A la proposición "Holanda logró el gol en el minuto 43 del segundo tiempo"; B la proposición "Argentina perdió el partido" y C "Argentina no salió campeón". Sean además A-mundos, B-mundos y C-mundos los mundos accesibles a *i* más próximos en los que tales proposiciones se verifican respectivamente. Puede haber muchos A-mundos no necesariamente de distinto grado de similaridad comparativa; incluso puede haber un

A-mundo en el que Argentina no haya perdido el partido. Pero bajo el supuesto de que es dificilísimo o casi imposible fácticamente que se pueda empatar o aun ganar un partido en sólo dos minutos, es mucho más similar a lo que ocurre en el mundo real *i*, que se verifique B y que por lo tanto la primera premisa A $\Box\!\!\rightarrow$ B sea verdadera en el A-mundo seleccionado. Los B-mundos también pueden ser muchos, inclusive puede existir un B-mundo en el que Holanda no haya logrado el gol. Para Lewis un tal B-mundo sería el más próximo a *i* y por lo tanto debería ser seleccionado para verificar la segunda premisa. Pero al hablante común no le interesa ese mundo, sino aquel B-mundo en el que además sea verdad que Holanda haya logrado el gol y que es el mundo que ha "fijado" el contexto contrafáctico. Bajo el supuesto de imposibilidad real de que Argentina perdiera ese partido y fuera campeón, se verifica también la segunda premisa B $\Box\!\!\rightarrow$ C. La conclusión se verifica sencillamente: puesto que en el B-mundo seleccionado sigue valiendo A, este B-mundo es un A-B-mundo, y puesto que en él también se verifica C, tal mundo es un A-B-C-mundo; por lo tanto es verdadero el contrafáctico A $\Box\!\!\rightarrow$ C.

El punto central de nuestra discrepancia con ambas teorías es que en ellas se exige la elección siempre del mundo antecedente más cercano posible respecto del mundo real *i*, sin tomar en cuenta la situación contrafáctica global o contexto contrafáctico. Con nuestra restricción, los mundos antecedentes seleccionados son cada vez más lejanos respecto de *i*. En nuestro ejemplo es obvio que un A-B-C-mundo sea menos similar a *i* que un A-mundo o un B-mundo, puesto que en el primero se verifican simultáneamente tres proposiciones falsas en *i*.

Puede observarse que con nuestra restricción, el contraejemplo de Lewis a la transitividad contrafáctica queda refutado. El mundo antecedente que debe seleccionarse para verificar la segunda premisa de la inferencia es ahora un mundo en el que Otto ha concurrido a la fiesta y Ana también. Y puesto que Waldo no quiere encontrarse con Otto,

entonces se falsifica el consecuente de la segunda premisa y se anula por lo tanto la posibilidad de que las premisas sean verdaderas y la conclusión falsa.

La cuestión que se nos presenta ahora es determinar si la aceptación de la transitividad contrafáctica no arrastra también la validez del refuerzo del antecedente, que habíamos convenido como adecuado rechazar. En otras palabras, ¿no estamos obligados a aceptar que partiendo de "A $\Box\!\!\rightarrow$ B" llegamos a "A.C $\Box\!\!\rightarrow$ B" mediante la ley lógica A.C $\Box\!\!\rightarrow$ A? Este argumento sería:

1) A.C $\Box\!\!\rightarrow$ A
2) A $\Box\!\!\rightarrow$ B
luego, 3) A.C $\Box\!\!\rightarrow$ B

Puesto que la primera premisa es una verdad lógica, ella es omitible en la deducción y por lo tanto la conclusión sería derivable de la segunda premisa solamente. Pero nuestro análisis de la importancia del contexto en la validación de un contrafáctico, muestra que en el caso de que A $\Box\!\!\rightarrow$ B fuera verdadero aisladamente, ello no prueba que sea verdadero también en el contexto en que apareció antes A.C $\Box\!\!\rightarrow$ A. En dicho contexto, el A-mundo tomado en cuenta para verificar A $\Box\!\!\rightarrow$ B debe respetar la restricción de ser un A-C-mundo y entonces puede ser que en ese A-C-mundo, la premisa A $\Box\!\!\rightarrow$ B resultara falsa, en cuyo caso también sería falsa la conclusión. Luego, si bien es cierto que de 1) y 2) por transitividad se sigue la conclusión, ello no indica nada acerca de lo que se infiere de la segunda premisa solamente. En síntesis, no puede ser 1) y 2) verdaderas y 3) falsa, pero sí puede ser 2) verdadera y 3) falsa. Aunque lo pareciera, esto no resulta paradójico; simplemente 2) tomada aisladamente tiene condiciones de verdad distintas a las que toma en el contexto donde interviene junto a 1).

Además, es posible mostrar también que la restricción impuesta a la selección de mundos antecedentes de un mismo contexto contrafáctico, no afecta en nada la invalidez del re-

fuerzo del antecedente. En efecto, si A $\Box\!\!\rightarrow$ B es verdadero, entonces en el A-mundo seleccionado B también es verdadero. Pero podría ocurrir que en alguno de los A-C-mundos que hay que seleccionar para validar A.C $\Box\!\!\rightarrow$ B, el consecuente B no se cumpliera.

Finalmente y a modo de síntesis, concluiremos enfatizando los aspectos que, a nuestro entender, surgen como los más importantes del análisis de las teorías que hemos realizado. En primer lugar, la teoría de Lewis se presenta con un grado de elaboración mucho mayor que la teoría de Stalnaker, permitiendo de ese modo un examen mucho más detallado, profundo y sutil de la lógica de los condicionales contrafácticos. En segundo lugar, recapitulemos los inconvenientes mayores presentados: a) para determinado tipo de condicionales contrafácticos —como los citados por Jackson— parece más relevante discutir cuestiones de causalidad que cuestiones relacionadas con la similaridad de los mundos posibles; b) aún para los casos de contrafácticos en los que resulta exitoso el análisis a través de las nociones de similaridad comparativa entre mundos posibles, las condiciones de verdad para ellos deben analizarse en forma contextual, a menos que nos resignemos a perder una de las inferencias más intuitivas del pensamiento natural como lo es la transitividad contrafáctica; y c) al uniformarse el tratamiento de los condicionales contrafácticos con antecedente lógicamente imposible y de los de antecedentes físicamente imposibles, se traiciona aparentemente el sentido de los condicionales contrafácticos contralegales.

SUMMARY

In modern logic there are many papers on the relevant role that counterfactual conditionals play in regard to the main problems in philosophy of science and with respect to their use in non-scientific contexts. These papers show, from differents points of view, the impossibility of a truth-functional analysis of counterfactual conditionals and the difficulty to precise their significance by means of truth conditions.

In this paper are discussed two theories about counterfactual conditionals and upon "possible words" and "comparative similarity" concepts. They are R. Stalnaker's theory in "A Theory of Conditionals" (in *Causation and Conditionals,* edited by E. Sosa, Oxford Univ. Press, 1975) and David Lewi's theory in *Counterfactuals* (Oxford, 1973). Especially, we study the truth conditions for counterfactual propositions and their logical consequences with respect to natural language.

Stalnaker's theory intends the same truth conditions for both conditional propositions and counterfactual propositions, because they deal with non-actual possible situations and therefore both conditionals are propositions about counterfactual worlds. He states the truth conditions for the conditional connective ">" as follows:

A > B is true in *i* if B is true in f(A,i)
A > B is false in *i* if B is false in f(A,i)

f is the selection function which takes a proposition and a possible world as arguments and a possible world as its value. This function selects for each antecedent A and world *j*, a particular possible world *i* in which A is true. Thus, the conditional sentence is true in the actual world when its consequent is true in the selected world. For selecting a world by the selection function, Stalnaker formulates four further conditions: (1) For any antecedent A and base world *i*, A must be true in f(A, i). This condition requires for the antecedent to be true in the selected world. (2) For any antecedent A and base world *i*, f(A, i) = λ if and only if there is not a possible world accesible to *i* in which A is true. This condition requires that the absurd world "λ" would be selected only when there is an antecedent that is impossible. Conditions (1) and (2) require that the world selected *differ minimally* from the actual world and this implies that there are no differences between the actual word and the selected world excepting those that are required implicity or explicity by the antecedent. (3) For all

base world *i* and all antecedents A, if A is true in *i*, then f(A, i) = i. This condition requires that the base world would be selected if it happens to be among the worlds in which the antecedent is true. (4) For all base world *i* and all antecedents B and B′, if B is true in f(B, i) and B′ is true in f(F, i), then f(B, i) = f(B′, i). This last condition ensures that the ordering among the worlds is established in a way that if any selection function establishes B as prior to B′ in the ordering, then no other selection function may establish B′ as prior to B. Conditions (3) and (4) together establish a total ordering of all selected world with the actual world preceding all of them.

Summarizing the most important objections to this theory:

1) The unified treatment of both types of conditionals (the indicative and the counterfactual) makes the truth value the same for any given pair of conditionals, when in fact one of them could be false as the example of Ernest Adams shows:

a) If Oswald did not kill Kennedy, then somebody else did it.

b) If Oswald had not killed Kennedy, then somebody else would have done it.

Obviously, the first conditional is true and the second may be false.

2) In Stalnaker's theory it is impossible (except in the vacuous case) to distinguish between "necessary" and "possible" counterfactuals:

a) If Oswald had not killed Kennedy, then somebody else would have done it.

b) If Oswald had not killed Kennedy, then somebody else might have done it.

Again, the first conditional may be false and the second may be true.

3) In Stalnaker's theory, the possible world selected by the selection function is a single world because it is supossed that there is only one world most similar to the actual world of all the worlds in which the antecedent holds. But, can a possible world differ from the actual world only with respect to what is implied by the antecedent and for the rest remain as it actually is? There are many ways in which, perhaps, the possible worlds can differ from the actual world, and nevertheless they can have the same degree of similarity with the actual world. If there are at least two possible worlds in which the antecedent holds with the same degree of similarity, the unicity of the selected world results in a problem.

David Lewis' theory is a very much elaborated and fruitful conception and therefore it overcomes the difficulties above mentioned, but it states others. In Lewis' theory there is not a single condicional

that can appear as an indicative or as a counterfactual conditional from the speaker's viewpoint about the truth value of the antecedent. There are two different types of conditionals and his theory only deals with counterfactual conditionals. Lewis introduces two counterfactual operators that pretend to embrace both counterfactuals "necessity" and "possibility", which were lost in Stalnaker's theory. They are the "would" operator "$\Box\!\!\rightarrow$", that must be read: "If it were the case that ———, then it would be the case that ———"; and the "might" operator "$\Diamond\!\!\rightarrow$", that must be read "If it were the case that ———, then it might be the case that ———". These operator are interdefinable and Lewis takes the "would" as primitive.

$$A \Box\!\!\rightarrow B = df \sim(A \Diamond\!\!\rightarrow \sim B)$$
$$A \Diamond\!\!\rightarrow B = df \sim(A \Box\!\!\rightarrow \sim B)$$

This theory is based on the semantic of possible worlds. Thus, there is a set of possible worlds one of which is the actual world. Each world *i* has a single sphere of accesibility *Si*, whose elements are the worlds that are accesible to *i* and differ from it just within certain limits. The accesible worlds more singular to *i* belong to inner spheres and those of less similarity belong to outer spheres. The set of all spheres of accesibility of a world *i*, forms a system of spheres $i around *i*, conforming a structure similar to Ptolemaic astronomy. The truth conditions of the "would" operator are as follow:

A $\Box\!\!\rightarrow$ B is true at a world *i* (according to a system of spheres $i) if and only if either

(1) no A–world belongs to any spheres S in $i, or

(2) some sphere S in $i does contain at least one A-world and A $\supset$ B holds at every world in S.

Condition (1) expresses the vacuous case. A counterfactual sentence is vacuously true if there is no antecedent-permitting sphere. Condition (2) gives the principal case: a counterfactual sentence is non-vacuously true if there is some antecedent-permitting sphere in which the the consequent holds at every antecedent world, and it is false otherwise.

There are some important points to consider:

1) According to the truth condictions in Lewis' theory, all counterfactual conditionals with impossible antecedents are true (vacuously true). It seems correct that some counterfactuals with impossible antecedents are true (namely counterfactuals with *logically impossible* antecedents). But other counterfactuals with impossible antecedents are false, as is the case of some counterfactuals whose antecedents deny what is affirmed by a physical law (*physically impossible*). For

example: "If Uranus and Neptune had not been submitted to gravitation, then Leverrier would have discovered Neptune from the irregularities of Uranus orbit";

2) In Stalnaker's theory the Principle of Excluded Middle Conditional is valid:

$$(A \Box\!\!\rightarrow B) \quad \vee \quad (A \Box\!\!\rightarrow \sim B)$$

In Lewis' theory this principle is not valid because both disjuncts migh be false. Nevertheless, Lewis affirms that the acceptation of this principle is plausible because in common language we cannot distinguish between the following expressions:

$$(1) \quad \sim(A \Box\!\!\rightarrow B) \qquad\qquad (2) \quad (A \Box\!\!\rightarrow \sim B)$$

Contrarily, if this principle is accepted, the difference between the "would" and the "might" counterfactuals is lost.

Nevertheless I think that the invalidity of the Principle of Excluded Middle Conditional is not as lamentable a consequence as Lewis has thought, because natural language sometimes distinguishes between expressions (1) and (2). If these expressions are interpreted in terms of Adam's example, it is observed that the first expression is true and the second is false. Moreover, with the same example both disjuncts of the Principle of Excluded Middle are false.

3) In Lewis' theory (likewise in Stalnaker's theory), the interference is not valid by strengthening the antecedent:

$$\frac{A \Box\!\!\rightarrow B}{\therefore (A.C) \Box\!\!\rightarrow B}$$

It is possible to form consistent sequences of counterfactuals and their negated opposites with stronger and stronger antecedents and alternative consequents between a sentence and its negations:

$$A \Box\!\!\rightarrow B \qquad (A.C) \Box\!\!\rightarrow \sim B$$
$$(A.C.D) \Box\!\!\rightarrow B \qquad (A.C.D.F) \Box\!\!\rightarrow \sim B \text{ etc.}$$

This is so because the counterfactuals are in Lewis' theory a variable strict conditional and not a strict conditional. Thus, the premise might be true and the conclusion false. There are many examples in natural language that are evidence of the fallacy of strengthening the antecedent, as it is done in the subjunctive and the indicative moods.

4) The inference of transitivity is considered by Lewis as a generalization of the fallacy of strengthening the antecedent and is therefore also invalid.

In opossition with the above case, natural language seems to observe the transitivity among counterfactual propositions. Michelson's experiment in order to prove the existence of ether provides us with a good example: "If ether had existed, then it would have presented an obstacle to light; and if ether had presented an obstacle, then the velocity of light would have experimented a drop. Therefore, if the ether had existed, then the velocity of light would have experimented a drop". Other more popular example is: "If Holland had scored a goal in the 43th minute of the second half, then it would have gained the match; and if Holland had gained the match then Argentine would not have been champion; therefore, if Holland had scored a goal in the 43th minute of the second half, then Argentine would not have been champion."

These and other examples show that both theories are far away from the use of counterfactual conditionals in natural languages. The invalidity of transitivity is caused by the fact that counterfactual operators are not strict conditionals; i.e., it is permitted to report different accesible worlds in the same counterfactual inference to validate the premises. So it is possible for the counterfactual $A \Box\!\!\rightarrow B$ and the counterfactual $B \Box\!\!\rightarrow C$ to be true and for the counterfactual $A \Box\!\!\rightarrow C$ to be false, because the B-world "selected" to validate $B \Box\!\!\rightarrow C$ would no be an A-world.

Contrarily, I argue that the truth conditions of counterfactuals are not independent from each other in the same counterfactual context. I assume that the world selected for validating the second premise $B \Box\!\!\rightarrow C$ also must be an A-world. So this will be an A-world and also an A-B-world. Since in this A-B-world, C is true, then $A \Box\!\!\rightarrow C$ is also true and the inference by transitivity is not yet a fallacy.

Now a new problem is raised: how is it possible that the strengthening of the antecedent and the transitivity can be invalid? It is right that, by transitivity, $(A.C) \Box\!\!\rightarrow B$ follows from $(A.C) \Box\!\!\rightarrow A$ and $A \Box\!\!\rightarrow B$. But since $(A.C) \Box\!\!\rightarrow A$ is a logical law it can be omitted and then $(A.C) \Box\!\!\rightarrow B$ follows only from $A \Box\!\!\rightarrow B$. Nevertheless this argument is wrong accordingly with our restriction. It is true that it is impossible for $(A.C) \Box\!\!\rightarrow A$, $A \Box\!\!\rightarrow B$ and $(A.C) \Box\!\!\rightarrow B$ to have the truth values true-true-false. It is also true that $(A.C) \Box\!\!\rightarrow A$ cannot be false (it is always true). But it is possible for $A \Box\!\!\rightarrow B$ to be true and for $(A.C) \Box\!\!\rightarrow B$ to be false, because $A \Box\!\!\rightarrow B$ may change its truth-value when it changes from being isolated to a context in which $(A.C) \Box\!\!\rightarrow B$ figures.

[*G. P.*]