



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE FILOSOFÍA Y LETRAS

**Evaluación de condicionales contrafácticos: Perspectivas  
lógicas y filosóficas**

TESINA

QUE PARA OBTENER EL TÍTULO DE:

LICENCIADA EN FILOSOFÍA

Presenta:

*Mitzi Dzahy Islas Paredes*

Tutor:

Mtro. Gabriel Ramos García

CIUDAD UNIVERSITARIA, CIUDAD DE MÉXICO, MAYO 2018



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# Índice

<b>Agradecimientos</b>	<b>2</b>
<b>Introducción</b>	<b>5</b>
<b>1. Condicionales</b>	<b>7</b>
1.1. Condicional material . . . . .	8
1.2. Condicional estricto . . . . .	12
1.3. Condicional contrafáctico . . . . .	16
<b>2. La teoría Stalnaker-Lewis</b>	<b>17</b>
2.1. La teoría de Robert Stalnaker . . . . .	17
2.2. La teoría de David Lewis . . . . .	27
2.3. Análisis comparativo . . . . .	37
2.3.1. Similitudes entre las teorías . . . . .	37
2.3.2. Diferencias entre las teorías . . . . .	38
2.4. Conclusiones . . . . .	39
<b>3. La teoría de Eric Hiddleston</b>	<b>40</b>
3.1. Objeciones a la teoría Stalnaker-Lewis . . . . .	40
3.2. Desarrollo de la teoría de Hiddleston . . . . .	43
<b>Conclusiones</b>	<b>59</b>

# Agradecimientos

Una de las primeras cosas que noté al ingresar a la licenciatura es que no era nada de lo que yo esperaba. El comienzo fue difícil, más aún por mi falta de bagaje filosófico, y esto me hizo sentir más de una vez que estaba fuera de lugar, que quizá debí elegir otra licenciatura, sin embargo trabajé por sobreponerme a esos sentimientos y ser mejor en los estudios. Me falta mucho por aprender, pero me siento feliz de haber encontrado mi lugar.

A mis padres, quienes son el corazón de este trabajo, gracias por el apoyo que siempre me han brindado y por todo su amor y comprensión, me han dado un hogar, me han enseñado que con esfuerzo y dedicación puedo lograr lo que me proponga. A mi madre, quien es ejemplo de bondad y templanza, te agradezco estar conmigo incluso cuando todo parece imposible, siempre tienes una palabra de aliento e intentas hacer de mí una mejor persona, aunque a veces soy demasiado testaruda para notarlo. A mi padre, quien es ejemplo de fortaleza, te agradezco que siempre hayas creído en mí incluso cuando yo no lo hago, me apoyaste como nadie al querer emprender este camino, gracias por la confianza y el amor que siempre me has dado.

A mi hermano Agustín, quién es mi único hermano y sé que siempre puedo contar con su apoyo, te agradezco tus extrañas formas de demostrar cariño y que te preocupes por mí a tu manera, agradezco que seas mi hermano y saber que estamos el uno para el otro.

A mi tía Martha y mi tío Rodolfo, quienes siempre me han apoyado, les agradezco todo su cariño y que se preocupen por mí. Sé que puedo contar con ustedes y me lo han hecho saber incontables veces, este trabajo también es para ustedes con todo mi cariño.

A la familia Nacif Goddard, quienes han estado conmigo desde el inicio de mi vida, les agradezco que me hayan hecho sentir bienvenida en su casa y en su familia, ustedes forman parte importante de mi vida y son la familia que escogí tener.

Al Mtro. Gabriel Ramos García, quien dirigió este trabajo y lo hizo excepcionalmente. Te agradezco toda la ayuda que me diste para lograr la investigación, por el tiempo que le dedicaste, las correcciones y anotaciones, sin ti este trabajo no sería lo que es ahora. Sin embargo, no es lo único por lo que te estoy agradecida, fuiste parte importante de mi formación académica, de no ser por ti seguramente no habría podido comprender muchas cosas o quizá habría aban-

donado este camino. Agradezco poder decir que no sólo eres mi profesor y mi asesor, sino que también eres mi amigo, has estado conmigo en mis momentos más oscuros y también en los buenos, gracias por todo lo que has hecho por mí.

A Thomas, quien, a pesar de haber llegado al final de este proceso, se ha convertido en el motor que impulsa este trabajo. Te agradezco todo tu amor, apoyo y comprensión, eres la persona por la que quiero ser mejor cada día y me impulsa a seguir adelante. Agradezco que tu vida se volviera a cruzar con la mía y que estés a mi lado.

Al seminario de Bestias Proposicionales, Claudia, David, Gabriel, Julio y Rodolfo, por todo lo que hemos aprendido juntos, por todas aquellas sesiones en las que nos hemos apoyado y discutido, y que muchas de ellas estuvieron dedicadas a nutrir este trabajo. Agradezco poder decir que no sólo son mis compañeros, sino que son mis amigos, son los mejores amigos que alguien podría pedir.

A Mario, quien ha sido mi amigo por muchos años, agradezco que estés siempre a mi lado a pesar de todos los desacuerdos que llegamos a tener, eres un gran amigo y te quiero mucho.

Al Mtro. Arturo González Yáñez, por leer este trabajo, por sus comentarios y correcciones, que fueron de mucha ayuda. Agradezco todas sus enseñanzas, fueron sus clases las que me dieron la pauta para elegir este tema que me ha fascinado tanto.

Al Dr. Cristian Gutiérrez Ramírez, quien ha sido parte importante de mi formación académica, agradezco que te hayas tomado el tiempo de leer este trabajo y tus comentarios, me alegra haber tenido la oportunidad de trabajar contigo, gracias por todas tus enseñanzas.

Al proyecto PAPIIT IA401717 “Pluralismo y normatividad en lógica y matemáticas”, por el apoyo otorgado para la terminación de esta tesina.

Al Dr. Ricardo Mena Gallardo, por su disposición a leer esta tesina, le agradezco su tiempo y sus atenciones.

Al Dr. Jesús Jasso Méndez, quién me hizo favor de leer y aprobar este trabajo.

A todos los que de alguna manera contribuyeron con este trabajo, ya fuera haciendo comentarios, sosteniendo una discusión conmigo, o simplemente con palabras de aliento, les agradezco infinitamente su apoyo, comprensión y cariño.

# Introducción

El presente trabajo tiene como objetivo presentar dos tipos de teorías que han intentado dar cuenta de la evaluación de los condicionales contrafácticos. La primera de ellas es la que se conoce como la teoría Stalnaker-Lewis, la cual se basa en la semántica de mundos posibles; la segunda es la teoría expuesta por Eric Hiddleston basada en modelos causales.

El tema de los condicionales ha sido de gran importancia dentro de la filosofía por diversas razones: por una parte, en lógica el condicional se encuentra estrechamente relacionado con la noción de consecuencia lógica; por otra parte, los condicionales parecen modelar una estructura causal en la que el consecuente es causado por el antecedente, y la causalidad ha sido un tema de gran interés dentro de la filosofía. El interés por los condicionales contrafácticos se da por su uso en textos científicos, históricos y filosóficos, y cómo es que enunciaciones contrafácticas nos pueden ayudar a tener conocimiento sobre hechos no actuales.

En el primer capítulo se van a desarrollar los antecedentes y el surgimiento de la necesidad de hacer una teoría a partir de la cual podamos dar cuenta de las condiciones de verdad para los condicionales contrafácticos. Se comenzará por dar los primeros esbozos que se dieron para describir las condiciones de verdad de los condicionales, después pasaremos al desarrollo del condicional material y las limitaciones que este presenta, se revisará la propuesta del condicional estricto como una teoría general de los condicionales y de igual modo se señalará sus deficiencias; por último se introducirán los condicionales contrafácticos.

El segundo capítulo está dedicado a explicar la teoría Stalnaker-Lewis. En el primer apartado se desarrolla la teoría de Robert Stalnaker sobre los condicionales, presentada en su texto *A theory of conditionals*<sup>1</sup>; aunque los condicionales contrafácticos están entre sus intereses, su teoría se presenta como una teoría general de los condicionales. Stalnaker introduce una nueva conectiva a la que llama “corner” ( $>$ ) cuya evaluación será intensional ya que utiliza la semántica de mundos posibles. La propuesta consiste en tomar el mundo posible más cercano al mundo base en el que el antecedente sea verdadero, y dentro de ese mundo posible evaluar el condicional, sin embargo, una de las objeciones a esta teoría va a ser aquello que entendemos por el “mundo posible más cercano al mundo base”, ya que podría tratarse de más de uno.

---

<sup>1</sup>Stalnaker, Robert. “A Theory of Conditionals”. In N. Rescher ed., *Studies in Logical Theory*. Oxford: Basil Blackwell, 1968. p. 98-112.

El segundo apartado de este capítulo trata sobre la propuesta presentada por David Lewis en su texto *Counterfactuals*<sup>2</sup>, donde postula una teoría para los contrafácticos basada en la teoría de Stalnaker, aunque con algunas diferencias. Lewis también utiliza la semántica de mundos posibles para la evaluación de dichos condicionales, pero en lugar de referirse al mundo posible más cercano, utiliza “esferas de accesibilidad” de los mundos posibles más cercanos. También distingue entre de dos tipos de condicionales contrafácticos, el contrafáctico “might” y el “would”.

Por último, en el tercer apartado del capítulo, se presenta un análisis comparativo de las teorías, las ventajas y desventajas que presentan una sobre otra. También se aborda si Lewis resuelve o no los problemas que se presentaron a la teoría de Stalnaker.

El tercer capítulo presenta una teoría alternativa para los condicionales contrafácticos que utiliza modelos causales, esta teoría es presentada por Eric Hiddleston en su texto *A causal theory of counterfactuals*<sup>3</sup>. Hiddleston intenta desarrollar una teoría que aventaje a la propuesta por Stalnaker-Lewis. Según él, al apelar a las relaciones causales entre antecedente y consecuente (que es algo que no pasa en las teorías semánticas de mundos posibles) la evaluación de los contrafácticos resulta —de alguna manera— más intuitiva<sup>4</sup>. El tratamiento de Hiddleston consiste en exponer los casos problemáticos de la teoría Stalnaker-Lewis, para después mostrar los elementos que van a conformar su teoría.

Al final de este trabajo se dará una breve comparación entre las teorías presentadas, las ventajas o desventajas que se pueden presentar y si es que pueden resultar al menos medianamente satisfactorias para decir que tenemos un método de evaluación de los condicionales contrafácticos.

---

<sup>2</sup>Lewis, David. *Counterfactuals*. Cambridge: Harvard University Press, 1973.

<sup>3</sup>Hiddleston, Eric. “A causal theory of counterfactuals”. *Noûs*, Vol. 39, No. 4 (Dec., 2005), p. 632-657.

<sup>4</sup>Lo que parece que Hiddleston está proponiendo es que podemos tener una evaluación de los contrafácticos que sea más cercana a cómo se entiende un contrafáctico en el habla cotidiana.



# 1. Condicionales

La problemática respecto al análisis semántico de las oraciones condicionales, como veremos a continuación, ha sido tema de discusión a lo largo de la historia. Dicho problema surge al preguntarse ¿qué es lo que se quiere decir cuando se utiliza una oración condicional? Y con base en qué decimos que una oración condicional es aceptable o no.

Algunas de las primeras definiciones del condicional de las que tenemos testimonio, son las de Filón, Diódoro y Crisipo presentadas por Sexto Empírico en sus *Esbozos Pirrónicos*<sup>5</sup>. Para que un condicional sea verdadero, de acuerdo con la definición de Filón, debe cumplirse que:

*”No concluye en una cosa falsa, partiendo de una verdadera.”*<sup>6</sup>

Es decir, un condicional es verdadero cuando su antecedente es falso o su consecuente verdadero; la definición de Diódoro dice:

*“La que ni podría ni puede concluir en una cosa falsa, partiendo de una verdadera.”*<sup>7</sup>

Según esta definición, para decir que un condicional es verdadero, este debe resultar verdadero en todo tiempo, lo cual es una afirmación más fuerte de la presentada por Filón ya que, en esa, basta que el antecedente sea falso o el consecuente sea verdadero para que un condicional sea verdadero, en tanto que en el caso de Diódoro se agrega el elemento de la verdad en todo tiempo; por su parte, en la definición de Crisipo<sup>8</sup>, un condicional es verdadero cuando:

*“La negación de lo que en él está como consecuente, se contrapone también a lo que en él está como antecedente.”*<sup>9</sup>

Así, un condicional es verdadero cuando la negación del consecuente es incompatible con el antecedente, es decir que genera una contradicción, lo cual exige más que las propuestas anteriores.

Además de estas propuestas, se han dado otras que intentan esclarecer los usos de las sentencias condicionales y que han tenido mayor relevancia en la historia de la filosofía, entre

---

<sup>5</sup>Sexto Empírico, *Esbozos pirrónicos*. Gredos, Madrid, 1993. P. 126

<sup>6</sup>Ibíd.

<sup>7</sup>Ibíd.

<sup>8</sup>En el texto de Sexto Empírico, se presenta esta definición como la dada por los que introducen la coherencia.

<sup>9</sup>Op cit.

ellos se encuentran el condicional material, el condicional estricto y el condicional contrafáctico, los cuales se desarrollarán en los siguientes apartados.

## 1.1. Condicional material

Con la formalización de la lógica<sup>10</sup> surge una forma de evaluación del condicional, la cual se conoce como condicional material y es representado con la conectiva “ $\supset$ ” así, “ $P \supset Q$ ” se lee como “Si  $P$  entonces  $Q$ ” y se define de la siguiente manera:

P	$\supset$	Q
1	1	1
1	0	0
0	1	1
0	1	0

Tabla 1: Condicional material.

Es decir, el condicional es verdadero cuando el antecedente es falso o el consecuente verdadero; o bien, el condicional sólo es falso cuando el antecedente es verdadero y el consecuente falso. Bajo esta definición, el condicional material se define de forma parecida a como Filón lo había definido.

Algunas de las propiedades formales que tiene el condicional material son las siguientes:

- Distributividad:  $(P \supset (Q \supset R)) \supset ((P \supset Q) \supset (P \supset R))$
- Transitividad:  $(P \supset Q) \supset ((Q \supset R) \supset (P \supset R))$
- Identidad:  $P \supset P$
- Permutación:  $(P \supset (Q \supset R)) \supset (Q \supset (P \supset R))$

Cada una de estas propiedades nos permite tener ciertas inferencias respecto al condicional, veamos cómo es que funciona cada una de ellas. La distributividad nos permite inferir que cada vez que tengamos un condicional cuyo consecuente es otro condicional, si el antecedente del primero implica el antecedente del segundo, entonces el antecedente del primero implicará al consecuente del segundo en todos los casos, tal y como lo muestra su tabla de verdad:

---

<sup>10</sup>La lógica clásica como se conoce hoy en día surgió con la publicación hecha por Gottlob Frege de *Conceptografía*, en 1879, en la cual aparece el primer sistema completo de lógica elemental.

(P	⊃	(Q	⊃	R))	⊃	((P	⊃	Q)	⊃	(P	⊃	R))
1	1	1	1	1	<b>1</b>	1	1	1	1	1	1	1
1	0	1	0	0	<b>1</b>	1	1	1	0	1	0	0
1	1	0	1	1	<b>1</b>	1	0	0	1	1	1	1
1	1	0	1	0	<b>1</b>	1	0	0	1	1	0	0
0	1	1	1	1	<b>1</b>	0	1	1	1	0	1	1
0	1	1	0	0	<b>1</b>	0	1	1	1	0	1	0
0	1	0	1	1	<b>1</b>	0	1	0	1	0	1	1
0	1	0	1	0	<b>1</b>	0	1	0	1	0	1	0

Tabla 2: Distributividad.

El condicional principal es una tautología, por lo que no importa cuáles sean los valores de verdad de P, Q o R siempre será posible hacer esa inferencia.

La transitividad nos permite “eliminar el término medio”, es decir si dado un condicional, entonces si tenemos otro condicional cuyo antecedente empata con el consecuente del primer condicional, entonces podemos pasar directamente del antecedente del primer condicional al consecuente del segundo, o dicho de otra forma, dados dos condicionales que compartan el consecuente de una y el antecedente de la otra, se puede ir directamente del antecedente de la primera al consecuente de la segunda<sup>11</sup> y queda evidenciado en su tabla de verdad:

(P	⊃	Q)	⊃	((Q	⊃	R)	⊃	(P	⊃	R))
1	1	1	<b>1</b>	1	1	1	1	1	1	1
1	1	1	<b>1</b>	1	0	0	1	1	0	0
1	0	0	<b>1</b>	0	1	1	1	1	1	1
1	0	0	<b>1</b>	0	1	0	0	1	0	0
0	1	1	<b>1</b>	1	1	1	1	0	1	1
0	1	1	<b>1</b>	1	0	0	1	0	1	0
0	1	0	<b>1</b>	0	1	1	1	0	1	1
0	1	0	<b>1</b>	0	1	0	1	0	1	0

Tabla 3: Transitividad.

La Identidad es una propiedad que permite que cualquier cosa sea implicada por sí misma, lo cual resultará siempre verdadero:

<sup>11</sup>Por exportación:  $(P \supset Q) \supset ((Q \supset R) \supset (P \supset R)) \equiv ((P \supset Q) \wedge (Q \supset R)) \supset (P \supset R)$

(P	$\supset$	P)
1	1	1
0	1	0

Tabla 4: Identidad.

Por su parte, la permutación permite cambiar el orden de los antecedentes cuando tenemos un condicional cuyo consecuente es otro condicional, así el antecedente del condicional principal pasa a ser el consecuente del condicional que se encuentra en el consecuente y el antecedente que se encontraba en ese condicional pasa a ser el antecedente del condicional principal:

(P	$\supset$	(Q	$\supset$	R))	$\supset$	(Q	$\supset$	(P	$\supset$	R))
1	1	1	1	1	1	1	1	1	1	1
1	0	1	0	0	1	1	0	1	0	0
1	1	0	1	1	1	0	1	1	1	1
1	1	0	1	0	1	0	1	1	0	0
0	1	1	1	1	1	1	1	0	1	1
0	1	1	0	0	1	1	1	0	1	0
0	1	0	1	1	1	0	1	0	1	1
0	1	0	1	0	1	0	1	0	1	0

Tabla 5: Permutación.

Aunque pudiera parecer que el condicional material ofrece una definición estándar de una estructura condicional, esto no es del todo cierto. El condicional material no es capaz de modelar un condicional entendido en el sentido cotidiano. Uno de los primeros en dar cuenta de esto fue C. I. Lewis <sup>12</sup>, quien criticó la interpretación del condicional expuesta por Russell y Whitehead en su texto *Principia Mathematica*<sup>13</sup>. Para mostrar cómo es que el condicional material no es adecuado para representar el condicional entendido de manera ordinaria propuso las llamadas paradojas del condicional material. Estas son fórmulas que son consideradas por la lógica proposicional como verdades lógicas, pero resultan controversiales para el sentido común cuando son representadas en el lenguaje natural.

Algunas de las fórmulas que dan pie a las paradojas de la implicación material son las siguientes:

<sup>12</sup>Véase: Lewis, C.I. *Implication and the algebra of logic*. En *Mind*, Vol. 21, No. 84 (Oct., 1912), p. 522-531.; Lewis, C. I., *Survey of symbolic logic*, University of California press, Berkeley, 1918. Y Lewis, C. I. & Langford C. H., *Symbolic Logic*, New York, 1932.

<sup>13</sup>Russell B. & Whitehead, A.N. *Principia mathematica*, Cambridge, 1910-1913, 3 vols.

$$(1) P \supset (Q \supset P)$$

$$(2) \neg P \supset (P \supset Q)^{14}$$

$$(3) (P \supset Q) \vee (Q \supset P)$$

En su texto *Implicación material*<sup>15</sup>, Raymundo Morado nombra a (1) como la paradoja positiva, a (2) como la paradoja negativa y a (3) como la paradoja de conmensurabilidad. Lo que cada una de estas fórmulas nos dice es que (1) una proposición verdadera puede ser implicada por cualquier proposición, (2) una proposición falsa puede implicar a cualquier proposición, (3) Cualesquiera dos proposiciones una implica a la otra o viceversa. Ahora revisaremos cómo es que estas fórmulas resultan controversiales al transformarlas al lenguaje natural.

Pensemos en las siguientes proposiciones: (a) “El pasto es verde”, (b) “Los unicornios existen” y (c) “el martes fue un día caluroso”; tenemos por una parte que (a) es una proposición verdadera y (b) es una proposición falsa, las cuales utilizaremos para sustituir en (1), (2) y (3). Si en (1) sustituimos a P por (a) y a Q por (b) tendríamos que “Si el pasto es verde entonces, si los unicornios existen entonces el pasto es verde”. Como sabemos que “El pasto es verde” es verdadera, nos da como resultado que “Si los unicornios existen entonces el pasto es verde” también sea verdadero, lo cual resulta poco intuitivo, no todos estarían dispuestos a aceptar ese condicional como verdadero por la simple verdad de consecuente. Si en (2) sustituimos a P por (b) y a Q por (c) tendríamos que “Si los unicornios no existen entonces, si los unicornios existen entonces el martes fue un día caluroso”. Dado que “Los unicornios existen” es falso, se sigue que “Si los unicornios existen entonces el martes fue un día caluroso” sea verdadero, que tampoco resulta intuitivo, no se aceptaría la verdad del condicional únicamente por la falsedad del antecedente. Y si en (3) sustituimos P por (a) y Q por (c) tenemos que “o bien si el pasto es verde entonces el martes fue un día caluroso, o bien si el martes fue un día caluroso el pasto es verde” es verdadero, pero no es intuitivo cómo de dos proposiciones alguna de ellas debe implicar a otra.

Las llamadas paradojas de la implicación material son algunas de las motivaciones que se tuvieron para descartar que el condicional material modelara las estructuras condicionales utilizadas en el lenguaje natural.

---

<sup>14</sup>Por exportación tenemos que  $(\neg P \wedge P) \supset Q$ , que es cuando se dice que “de una contradicción se sigue lo que sea”.

<sup>15</sup>Morado, Raymundo. Instituto de Investigaciones Filosóficas, UNAM. 20 de Marzo de 2018 “Implicación Material”, <http://www.filosoficas.unam.mx/~morado/Cursos/11FilLog3Sem/110901.pdf>

## 1.2. Condicional estricto

El objetivo de C.I. Lewis es crear una lógica libre de paradojas y que recoja el sentido estricto del condicional del lenguaje ordinario, por lo que crea su propio sistema del condicional estricto y que se distingue del de Russell el cual es un sistema basado en el condicional material<sup>16</sup>. La característica del condicional estricto es que se vale de conceptos modales para su análisis, así: A implica estrictamente B ( $A \prec B$ ), es verdadero cuando hay una relación de necesidad entre antecedente y consecuente, (es imposible que el antecedente sea verdadero y el consecuente falso).

Las motivaciones de Lewis están basadas en el problema de la deducción y que la implicación propuesta por Russell es incapaz de caracterizarla, ya que en la deducción lo que se busca es establecer la relación entre antecedente y consecuente, la cual, es independiente de los valores de verdad. En su texto *Implication and the algebra of logic*<sup>17</sup>, Lewis hace mención de la definición del condicional material como:  $[(P \supset Q) = (\neg P \vee Q)]$  Df.] es decir, ‘P implica Q’ es definido como: ‘o P es falso o Q es verdadero’; pero hacer uso de una disyunción puede referir al menos a dos tipos: uno extensional y otro intensional. Decir que una disyunción es extensional significa que esta depende de los valores de verdad de las proposiciones conectadas, es decir, es veritativo-funcional; se dice que una disyunción es intensional cuando, a pesar de poder saber si es verdadera, es problemático decidir cuál de sus partes es verdadera. Desde el punto de vista de Lewis, el significado de ‘implica’ en el lenguaje ordinario refiere a la relación de deducción y esta es intensional ya que su verdad está vinculada a algo más que sólo los valores de verdad; Lewis busca que su sistema sea capaz de dar cuenta tanto de las relaciones extensionales como de las intensionales.

La propuesta de Lewis consta del desarrollo de un sistema formal que comienza a desarrollar en su texto *Survey of symbolic logic*<sup>18</sup> y que después modifica y complementa en su texto *Symbolic Logic*<sup>19</sup>, que hace en colaboración con C. H. Langford y tiene como símbolo primitivo a  $\prec$ . El sistema se puede definir como sigue:

---

<sup>16</sup>El interés de Russell era tener un significado lógico de las expresiones condicionales que estuviera restringido al contexto de las pruebas matemáticas, mientras que Lewis buscaba rescatar el significado de las expresiones condicionales de manera más amplia.

<sup>17</sup>Lewis, C. I. *Implication and the algebra of logic*. En *Mind*, Vol. 21, No. 84, 1912. p. 522-531.

<sup>18</sup>Lewis, C.I. *Survey of symbolic logic*, Univ. of California, 1918.

<sup>19</sup>Lewis, C.I. & Langford, C.H. *Symbolic Logic*, New York, 1932.

Nociones primitivas:

- Proposiciones:  $(P, Q, R, \dots)$ .
- Negación:  $\neg P$  ( $P$  es falsa).
- Imposibilidad:  $\neg \diamond P$ <sup>20</sup> ( $P$  es imposible, o es imposible que  $P$  sea verdadera).
- Producto lógico:  $P \times Q$  o  $PQ$ . ( $P$  y  $Q$  los dos, o  $P$  es verdadero y  $Q$  es verdadero).
- Equivalencia:  $P=Q$ . (La definición de la relación).

Relaciones diádicas:

- I) Consistencia ( $\circ$ ):  $P \circ Q = \text{df. } \neg \neg \diamond (PQ)$  ie.  $\diamond(PQ)$ .
- II) Implicación estricta ( $\prec$ ):  $P \prec Q = \text{df. } \neg \diamond(P \neg Q)$ .
- III) Implicación material ( $\supset$ ):  $P \supset Q = \text{df. } \neg(P \neg Q)$ .
- IV) Suma lógica estricta ( $\wedge$ ):  $P \wedge Q = \text{df. } \neg \diamond(\neg P \neg Q)$
- v) Suma lógica material ( $+$ ):  $P + Q = \text{df. } \neg(\neg P \neg Q)$
- VI) Equivalencia estricta ( $=$ ):  $P = Q = \text{df. } (P \prec Q)(Q \prec P)$
- VII) Equivalencia material ( $\equiv$ ):  $P \equiv Q = \text{df. } (P \supset Q)(Q \supset P)$

Lewis divide las relaciones diádicas (las siete que se acaban de presentar más la relación primitiva de producto lógico) en dos grupos: las relaciones materiales ( $PQ$ ,  $P \supset Q$ ,  $P + Q$ ,  $P \equiv Q$ ) y las relaciones estrictas ( $P \circ Q$ ,  $P \prec Q$ ,  $P \wedge Q$ ,  $P = Q$ ); las primeras se encuentran en cualquier cálculo de implicación material mientras que las segundas entrañan la idea de imposibilidad, la cual no se encuentra en los cálculos de implicación material.

Postulados:

1.  $PQ \prec QP$ . Si tanto  $P$  como  $Q$  son ambos verdaderos, implica estrictamente que  $Q$  y  $P$  son ambos verdaderos.
2.  $QP \prec P$ . Si tanto  $Q$  como  $P$  son ambos verdaderos, implica estrictamente que  $P$  es verdadero.

---

<sup>20</sup>En *Survey of symbolic logic*, Lewis lo presenta como  $\sim P$ , para evitar confusiones con la negación se representará como en *Symbolic logic* con ayuda del símbolo  $\diamond$ .

3.  $P \rightarrow P$ . Que P es verdadero implica estrictamente que P es verdadero y P es verdadero.
4.  $P(QR) \rightarrow (PQ)R$ . Si P es verdadero y tanto Q como R son ambos verdaderos, implica estrictamente que tanto P como Q son ambos verdaderos, y R es verdadero.
5.  $P \rightarrow \neg(\neg P)$ . Que P sea verdadero, implica estrictamente que es falso que P sea falso.
6.  $(P \rightarrow Q).(Q \rightarrow R) \rightarrow (P \rightarrow R)$ . Si P implica estrictamente Q y Q implica estrictamente R, implica estrictamente que P implica estrictamente R.
7.  $P \rightarrow P \rightarrow Q \rightarrow Q$ . Si tanto P como P implica estrictamente Q son verdaderos, implica estrictamente que Q es verdadero.
8.  $\diamond(PQ) \rightarrow \diamond P$ . Si es posible que tanto P como Q sean ambas verdaderas, entonces implican estrictamente que es posible que P sea verdadera.
9.  $(\exists P, Q): \neg(P \rightarrow Q) \rightarrow \neg(P \rightarrow \neg Q)$ . Existen P y Q tales que no es el caso que P implique estrictamente Q, ni que P implique estrictamente  $\neg Q$ .

Operaciones:

- a) Sustitución: Cualquier proposición puede ser sustituida por P, Q, R, . . . , etc. Cualquier par de expresiones que estén relacionadas por “=”, pueden ser sustituidas una por la otra.
- b) Inferencia: Si P es afirmado y  $P \rightarrow Q$  es afirmado, entonces Q puede ser afirmado.
- c) Producción: Si P y Q son afirmados separadamente, entonces PQ puede ser afirmado.

Como ya se ha apuntado, existe relación con los teoremas válidos para la implicación material, esto se debe a que todo lo que sea válido para la implicación estricta lo es también para la implicación material, pero no al revés, es decir:  $[(P \rightarrow Q) \rightarrow (P \supset Q)]$  y  $\neg[(P \supset Q) \rightarrow (P \rightarrow Q)]$ , ya que hay teoremas válidos para la implicación material que no es posible demostrar para la implicación estricta, por ejemplo: de que no pueda darse la verdad de P con la falsedad de Q,  $\neg(P \rightarrow Q)$ , no implica necesariamente que de P se deduzca Q,  $(P \rightarrow Q)$ , esto se debe a que  $\neg(P \rightarrow Q)$  es equivalente (en los sistemas de implicación material) a decir que  $P \supset Q$  y de eso no podemos llegar a  $P \rightarrow Q$  dado que hace falta el aspecto modal de decir que de hecho es imposible que ocurra  $\neg(P \rightarrow Q)$ .

Dentro de las relaciones diádicas, podemos observar que Lewis tiene un símbolo para la consistencia (o), para la cual también tenemos algunos teoremas destacados:



- I)  $(PQ) \prec (PoQ)$ . Si dos proposiciones son ambas verdaderas, entonces son consistentes.
- II)  $P \prec Q = \neg(Po\neg Q)$ . Si una proposición se deduce de otra, no es consistente la verdad de la primera con la falsedad de la segunda.
- III)  $\neg(Po\neg P): (P \prec R) (Q \prec S) (PoQ): \prec (RoS)$ . Un sistema deducido de postulados consistentes, es todo él consistente.
- IV)  $(P \prec Q) (P \prec \neg Q) : \prec \neg(PoP)$ . Si de una proposición se sigue otra y también se sigue su negación, la primera es inconsistente.
- V)  $\diamond P \prec \diamond(P \vee Q)$ . Si una proposición es posible, entonces es posible que esa o cualquier otra proposición.
- VI)  $\neg(PoP) \prec \neg(PoQ)$ . Si una proposición es inconsistente, no puede ser consistente con ninguna otra.
- VII)  $\diamond(PQ) \prec (\diamond P \diamond Q)$ . Si es posible que dos proposiciones sean ambas verdaderas, entonces es posible que la primera sea verdadera y es posible que la segunda también sea verdadera.
- VIII)  $\neg \diamond P \prec \neg \diamond(PQ)$ . Si una proposición no es posible, entonces no es posible que sea verdadera con otra proposición.
- IX)  $\neg \diamond P .:= \neg(PoQ). \neg(Po\neg Q) .:= (P \prec Q) . (P \prec \neg Q)$ . Si una proposición es imposible, equivale estrictamente a que es inconsistente con cualquier proposición la negación de dicha proposición, lo que equivale a decir que de la proposición imposible se deduce cualquier proposición y su negación.
- X)  $\neg \diamond \neg P .:= (Q \prec P).(\neg Q \prec P)$ . Que una proposición sea imposible de ser falsa, equivale a decir que puede ser deducida de cualquier otra, así como de su negación.

A pesar de los esfuerzos de Lewis por formar un sistema libre de paradojas, el sistema no es capaz de modelar el sentido estricto del condicional y es posible desarrollar lo que se llaman las paradojas de la implicación estricta (teoremas IX y X), i.e. de una proposición imposible, se puede deducir cualquier otra proposición,  $\neg \diamond P \prec (P \prec Q)$ ; y de una proposición necesaria se deduce que esta es implicada por cualquier otra proposición,  $\square P \prec (Q \prec P)$ . Como se puede observar, las paradojas de la implicación estricta tienen una estructura similar a las paradojas de la implicación material, pero son aún más fuertes ya que tienen el peso de la modalidad.

### 1.3. Condicional contrafáctico

Los condicionales contrafácticos son condicionales expresados en modo subjuntivo, no obstante, es necesario hacer precisiones: si es verdad que todos los condicionales contrafácticos se expresan en modo subjuntivo, no es cierto que todos los condicionales expresados en modo subjuntivo son contrafácticos. El modo subjuntivo es un modo verbal que hace referencia a una acción hipotética o posible ya sea del presente o del pasado, por lo que no hay nada que nos impida formular un condicional en modo subjuntivo el cual carezca del carácter contrafáctico: “Si la UNAM abriera más lugares, habría más jóvenes con oportunidades”; el condicional está expresado en modo subjuntivo, ya que cumple con expresar una situación posible o hipotética, sin embargo, la UNAM en algún momento posterior podría ser capaz de abrir más lugares y cumplirse la situación hipotética que se enuncia, por lo que carece del carácter contrafáctico.

El tiempo en el que los condicionales contrafácticos están formulados es en pretérito pluscuamperfecto del subjuntivo el cual corresponde a las acciones hipotéticas o posibles no realizadas, situadas en el pasado, es decir, acciones que podrían haber tenido lugar en el pasado en otras circunstancias diferentes a las actuales, lo que en el condicional nos garantiza la falsedad del antecedente.

El interés por dicho condicional comenzó cuando los filósofos positivistas se dieron cuenta que los términos disposicionales<sup>21</sup> que eran utilizados todo el tiempo en los textos científicos y filosóficos podían ser expuestos en estructuras contrafácticas que, al ser tratados como condicionales materiales, serían trivialmente verdaderos<sup>22</sup>, pero aceptar que la semántica de dichos condicionales no era veritativo-funcional constituía un problema al proyecto positivista.

Algunas de las teorías para evaluar los condicionales contrafácticos son la basada en los mundos posibles y la que se basa en modelos causales. Las teorías representativas del enfoque de los mundos posibles son las expuestas por Robert Stalnaker y David Lewis; y la teoría de modelos causales, la que utilizaremos para este trabajo será la expuesta por Eric Hiddleston.

---

<sup>21</sup>Aquellos términos que se refieren a las disposiciones de objetos o sistemas.

<sup>22</sup>El antecedente del condicional contrafáctico es siempre falso y el condicional material es verdadero cuando el antecedente es falso o el consecuente verdadero.

## 2. La teoría Stalnaker-Lewis

### 2.1. La teoría de Robert Stalnaker

Stalnaker expone su teoría sobre los condicionales en su texto de 1968, *A theory of conditionals*<sup>23</sup>, la cual atiende tres de los problemas que suelen presentarse en la evaluación de los condicionales; el primero de ellos y el principal es lo que se conoce como el problema lógico de los condicionales, el cual consiste en describir las propiedades formales de una función condicional; el segundo es el problema pragmático de los contrafácticos que atiende la idea de que las propiedades formales de las funciones condicionales no bastan para determinar el valor de verdad de los contrafácticos y por lo tanto necesitan de una evaluación distinta; el tercer problema es el epistemológico que puede ser descrito de la siguiente forma: dado que algunos condicionales contrafácticos parecen ser afirmaciones contingentes de hechos no realizados y que se supone que las aseveraciones contingentes tienen una verificación empírica, ¿cómo es posible que un condicional sea tanto empírico como contrafáctico?

Para resolver estos problemas, Stalnaker presenta una teoría que intenta dar cuenta de las construcciones condicionales en general, para ello introduce por una parte un sistema formal, y por otra un aparato semántico. Comenzaremos por analizar las motivaciones y la propuesta de su aparato semántico.

Stalnaker aborda el problema partiendo de las siguientes cuestiones: (i) cómo es que deliberamos si una afirmación condicional es verdadera o falsa y (ii) qué elementos lógicos son relevantes para determinar su verdad. Para contestar a lo anterior, consideraremos tres posibles respuestas: 1) Examinar las sentencias condicionales con el análisis de funciones de verdad a partir de un test de verdad o falsedad: 2) Existe una idea de conexión entre el antecedente y el consecuente de los condicionales, por ello tienen que ser entendidos como afirmaciones que tienen una conexión causal o lógica entre el antecedente y el consecuente; 3) El test de Ramsey<sup>24</sup> que consiste en añadir de manera hipotética el antecedente a nuestro conjunto de creencias para considerar si el consecuente es o no verdadero, si el consecuente es verdadero, el condicional es añadido al conjunto de creencias, mientras que si es falso se descarta.

---

<sup>23</sup>Stalnaker, Robert. "A Theory of Conditionals". In N. Rescher ed., *Studies in Logical*.

<sup>24</sup>En: Ramsey, "General propositions and causality", en su colección de artículos: *Foundations*, Londres, 1978. P. 143.

Pero ninguna de estas respuestas resulta completamente satisfactoria, 1) Atiende a la estructura del condicional material, la cual no siempre es posible aplicar a todas las estructuras condicionales, debido a que la falsedad del antecedente no siempre es razón suficiente para aceptar un condicional como verdadero; 2) Es una propuesta que intenta resolver el problema que existe en 1) atendiendo a la idea de conexión entre antecedente y consecuente, así la evaluación no consiste en los valores de verdad de las partes del condicional, sino en la relación de dichas partes, es decir si la conexión se sostiene el condicional es verdadero mientras que si no se sostiene es falso, pero la idea de conexión por sí misma es ambigua y oscura, parece que la presencia de una conexión entre las partes no es algo necesario en los enunciados condicionales; en 3) la conexión será a veces relevante y a veces no, dependiendo de si el sujeto cree que existe o no una relación entre antecedente y consecuente, sin embargo el test sólo toma en cuenta el caso en el que el sujeto no tiene una opinión acerca de la verdad o falsedad de las sentencias expresadas en el condicional, si intentáramos generalizarlo podríamos hacerlo de manera satisfactoria cuando se cree o se sabe que el antecedente es verdadero, pero no cubre el caso en el que el antecedente se cree o se sabe falso ya que no puede ser añadido al conjunto de creencias sin que se genere una contradicción.

Para enfrentar las dificultades que se presentan con el test de Ramsey necesitaríamos introducir hipotéticamente el antecedente (que creemos o sabemos falso) y hacer ajustes a nuestro conjunto de creencias para mantener la consistencia. Aceptar lo anterior nos hace preguntarnos la manera en la que dichos ajustes deben realizarse, que es justamente el problema pragmático de los condicionales, pero Stalnaker considera que es posible responder a la pregunta a partir del siguiente método de evaluación de condicionales:

*Primero, añadir el antecedente (hipotéticamente) al conjunto de creencias; segundo, hacer todos los ajustes requeridos para mantener la consistencia (sin modificar la creencia hipotética del antecedente); finalmente, considerar si el consecuente es verdadero o no.*<sup>25</sup>

Con este método se deja de lado el problema de inconsistencia que se presentaba cuando el antecedente era o se creía falso —si sólo lo utilizamos como una forma de encontrar condiciones de verdad—, pero para evaluar un condicional de esta manera es necesario no restringirse a un solo contexto particular de creencia ya que dicho contexto va a depender de los ajustes que sean requeridos.

---

<sup>25</sup>Op cit, P. 102.

Hasta aquí, se ha hablado de cuándo vamos a creer que un condicional es verdadero o falso, el siguiente problema que es importante enfrentar es la transición de pasar de las condiciones de creencia a las condiciones de verdad, ya que nos interesa saber cuándo un condicional es verdadero y no solamente cuándo un condicional se cree verdadero. Para ello se vale del concepto de *mundo posible*. Usando esta noción, Stalnaker se acerca más a su propuesta:

*Considere un mundo posible en el que A es verdadero, y que difiere mínimamente del mundo actual. “Si A, entonces B” es verdadero/falso sólo en el caso en el que B es verdadero/falso en ese mundo posible.*<sup>26</sup>

Este análisis de los condicionales proporciona un aparato sobre el cual es posible construir una teoría semántica formal. Stalnaker hace uso de la semántica del sistema de lógica modal propuesto por Saul Kripke<sup>27</sup>. El sistema propuesto por Stalnaker cuenta con una tripleta ordenada  $(K, R, \lambda)$  a la que llama M entendida de la siguiente manera: (1) Un conjunto K de mundos posibles; (2) una relación R de accesibilidad entre los mundos posibles del conjunto K; y (3) un mundo absurdo  $\lambda$  que no tiene acceso a ningún mundo ni es accesible para ningún otro mundo, es un mundo absurdo ya que en él todas las contradicciones y sus consecuencias lógicas son verdaderas, este mundo absurdo tiene como propósito permitir los casos en los que el antecedente de un condicional de la forma: “Si A entonces B” es imposible.

Es necesario añadir al sistema (4) una función de selección  $f$ , que tiene como argumentos proposiciones y mundos posibles, y como valor a un mundo posible, La función selecciona —para cada mundo posible y para cada antecedente A de una proposición condicional— un mundo posible en el que A es verdadero y el condicional será verdadero en el mundo base cuando el consecuente es verdadero en el mundo seleccionado por la función. A partir de estos elementos es posible establecer de manera formal las reglas semánticas para el condicional utilizando la conectiva “ $>$ ” (leído como *corner*) para representar un condicional: De una proposición A *entonces* B, decimos que es verdadera en un mundo  $w$  cuando el consecuente (B) es verdadero en el mundo seleccionado por la función  $f(A, w)$ ; el condicional será falso cuando el consecuente es falso en el mundo seleccionado por la función; es decir<sup>28</sup> (i)  $A > B$  es V en  $w$ , si B es V en  $f(A, w)$ ; (ii)  $A > B$  es F en  $w$ , si B es F en  $f(A, w)$ .

---

<sup>26</sup>Ibíd.

<sup>27</sup>Kripke, S. “Semantical analysis of modal logics”, I, *Zeitschrift für mathematische Logik and Grundlagen der Mathematik*, vol. 9, p. 67-96, 1963.

<sup>28</sup>Op cit. P. 103

La función-selección tiene como objetivo seleccionar el mundo posible más similar al mundo desde el cual se está partiendo (mundo base), con lo cual se precisa la noción de similitud entre mundos. Para ello, Stalnaker establece cuatro condiciones que debe cumplir la función-selección: a) Es necesario que, en el mundo seleccionado por la función, el antecedente del condicional sea verdadero, para ver qué ocurre con el consecuente y se pueda determinar el valor de verdad del condicional; b) Si no existe ningún mundo accesible desde el mundo base en el que el antecedente sea verdadero, es decir el antecedente es imposible, la función seleccionará el mundo absurdo  $\lambda$ ,  $f(A, w) = \lambda$ ; c) Si el antecedente del condicional es verdadero en el mundo base, la función-selección seleccionará al mundo base como el mundo más similar,  $f(A, w) = w$ ; d) Dados dos antecedentes, A y B, si A es verdadero en un mundo seleccionado donde B es verdadero, y B es verdadero en un mundo seleccionado donde A es verdadero, entonces el mundo seleccionado para ambos antecedentes debe ser el mismo  $f(A, w) = f(B, w)$ .

Las condiciones mencionadas son necesarias para que la interpretación pueda ser utilizada como una explicación del condicional, sin embargo, dichas condiciones semánticas no son suficientes, hace falta explicar en qué se basa la función selección y cómo es que se ordenan los mundos posibles. Para esto es que Stalnaker desarrolla un sistema formal.

El sistema desarrollado para los condicionales es coextensivo con el sistema formal C2<sup>29</sup>, que tiene como conectivas primitivas a  $\supset$  y  $\neg$ , se incluyen  $\vee$ ,  $\wedge$  y  $\equiv$  definidas de la manera usual y una conectiva condicional,  $>$  (corner). La manera en que se definen en términos de  $>$  los conceptos modales y condicionales es la siguiente:

- I)  $\Box A = \text{df. } \neg A > A$
- II)  $\Diamond A = \text{df. } \neg(A > \neg A)$
- III)  $A < > B = \text{df. } (A > B) \wedge (B > A)$ .

Las reglas de inferencia son *modus ponens* (Si A y  $A \supset B$  son teoremas, B es teorema) y *necesitación* (Si A es teorema,  $\Box A$  es teorema). Habiendo definido el lenguaje y establecido las reglas de inferencia, se añade un esquema de siete axiomas:

- a) Cualquier tautología que sea una fórmula bien formada es un axioma.

---

<sup>29</sup>El sistema C2 es presentado por Stalnaker en "A semantic analysis of conditional logic", y en ese mismo artículo prueba que C2 es un sistema completo.

- b)  $\Box(A \supset B) \supset (\Box A \supset \Box B)$
- c)  $\Box(A \supset B) \supset (A > B)$
- d)  $\Diamond A \supset ((A > B) \supset \neg(A > \neg B))$
- e)  $(A > (B \vee C)) \supset ((A > B) \vee (A > C))$
- f)  $(A > B) \supset (A \supset B)$
- g)  $(A < > B) \supset ((A > C) \supset (B > C))$

Bajo este sistema formal, su conectiva corner estaría situada entre el condicional estricto y la implicación material<sup>30</sup>, sin embargo, carece de ciertas propiedades que tienen ambos condicionales. Stalnaker presenta tres características de su conectiva condicional: (1) El condicional corner no es una conectiva transitiva, de  $A > B$  y  $B > C$  no es posible inferir  $A > C$ , pensemos en las siguientes premisas, “Si Juan hubiera salido con María, habría contraído una enfermedad de transmisión sexual” “Si Juan no hubiera tenido novia, habría salido con María”, por lo tanto “Si Juan no hubiera tenido novia, habría contraído una enfermedad de transmisión sexual”<sup>31</sup>. En el ejemplo anterior alguien podría afirmar las premisas, pero no la conclusión. Para hacer más claro lo anterior veamos otra regla de transitividad que no se cumple con el condicional corner, el refuerzo del antecedente: de  $A > B$  no podemos inferir que  $(A \wedge C) > B$ . Pensemos en las siguientes oraciones, “Si hubiera raspado un cerillo entonces se habría encendido” y “si hubiera raspado un cerillo y hubiera estado mojado, entonces se habría encendido”. (2) La negación del condicional también difiere, negar el condicional corner equivale a un condicional con el mismo antecedente pero consecuente opuesto (cuando el antecedente no es imposible):  $\Diamond A: \neg(A > B) \equiv (A > \neg B)$ , esto es bastante intuitivo cuando pensamos en contrafácticos. Contradecir un contrafáctico es tanto como contradecir el consecuente manteniendo el antecedente: Para negar “Si Colosio no hubiera sido asesinado, habría ganado las elecciones”, diríamos: “Si Colosio no hubiera sido asesinado, no habría ganado las elecciones”. (3) La inferencia de contrapositiva tampoco es válida para la conectiva corner, de  $A > B$  no se sigue que  $\neg B > \neg A$ , veamos el siguiente ejemplo: “Si me hubiera levantado temprano, habría llegado temprano a la cita”, “si no hubiera llegado temprano a la cita, no me habría levantado temprano.”.

Como se mencionó al principio, Stalnaker tiene como objetivo atender tres de los problemas

---

<sup>30</sup>Por los axiomas 3 y 6.

<sup>31</sup>La conectiva corner no es transitiva, sin embargo, la conectiva bicondicional ( $< >$ ) sí lo es, ya que el bicondicional es una relación de equivalencia y eso hace que sea simétrica y transitiva.

que se presentan en la evaluación de los condicionales y, una vez que ha quedado expuesta su teoría, es posible considerar la relación de dichos problemas con su teoría.

El primero de ellos es el problema lógico de los condicionales que atiende a la pregunta: ¿cuáles son las propiedades formales de los condicionales? La cual puede tener distintas respuestas. Una forma tradicional<sup>32</sup> de responder a la pregunta sería elegir un análisis que muestre que las oraciones ambiguas u objetables pueden eliminarse o remplazarse. Dicho análisis debe mostrar aquello que es subyacente al concepto o debe remplazar aquello que es vago en él. Una teoría que sostiene que el lenguaje ordinario es sistemáticamente erróneo sobre el concepto que se quiere explicar, o que se extravía en los diferentes sentidos que tienen los conceptos, pierde fuerza explicativa. Una respuesta a partir de la teoría semántica de los condicionales significa ver a la teoría como la construcción del concepto que remplaza la noción ambigua del lenguaje ordinario, o como una explicación del uso común del concepto, es decir, la teoría no va a explicar las reglas que operan sobre una lengua, sino que intenta explicar la estructura de un concepto, por ello no se debe entender que la teoría es una descripción de los usos lingüísticos. Stalnaker piensa que su teoría presenta una ventaja sobre la resolución de este primer problema, ya que da una interpretación general que evita dividir los sentidos de un concepto.

El segundo problema a tratar es el problema pragmático de los contrafácticos. Se ha dicho que las oraciones condicionales son susceptibles de ser ambiguas, dicha ambigüedad es producto de diferentes factores: las intenciones y/o creencias del hablante, el contexto en el que se enuncia una oración contrafáctica, etc. Para hablar de cómo es que las oraciones condicionales son ambiguas, es necesario distinguir tres tipos de ambigüedad: i) ambigüedad sintáctica: se presenta cuando una oración tiene más de una estructura gramatical; ii) ambigüedad semántica: cuando las oraciones/palabras tienen más de un significado; iii) ambigüedad pragmática: se presenta cuando la interpretación de una oración depende del contexto en que se enuncia. Stalnaker toma la ambigüedad de los condicionales como pragmática, por lo que el valor de verdad de una enunciación condicional va a depender de la especificación de la *s*-función. La razón de tratar la ambigüedad de los condicionales como una ambigüedad pragmática y no semántica, nos dice que es por simplicidad y coherencia sistemática<sup>33</sup>, para él, la semántica de un condicional tiene una estructura común para las oraciones condicionales que le proporciona a la conectiva un sentido único, mientras que la función-selección brinda el parámetro de inter-

---

<sup>32</sup>Stalnaker, Robert. "A Theory of Conditionals". In N. Rescher ed., *Studies in Logical Theory*. Oxford: Basil Blackwell, 1968. P. 98-112.

<sup>33</sup>*Ibid.* P. 109.



pretación.

Podemos entender perfectamente a alguien que usa una oración condicional sin especificar explícitamente una *s*-función, lo que nos lleva a que hay más reglas de las establecidas en la semántica del uso de condicionales, dichas reglas atienden a la pragmática de los condicionales. Para esto, Stalnaker desarrolla una pequeña investigación a partir de la cual se podría dar tratamiento a este problema<sup>34</sup>: (1) Una función lógica con conectiva condicional puede ser candidato a tener forma de ley natural. Las leyes naturales no nos dicen únicamente que todo lo que sea A es B, sino que, para cualquier C posible, si fuera A, entonces sería B. Las leyes entrañan a los contrafácticos. Para ilustrar esto veamos el siguiente ejemplo: “Todos los mamíferos son vertebrados” no es lógicamente equivalente a “Todos los no mamíferos son no invertebrados”, así que decir: “Si este animal hubiera sido mamífero, habría sido vertebrado” es una instancia de “Todos los mamíferos son vertebrados”; (2) Pensemos en el problema general de la proyectabilidad planteado por Goodman<sup>35</sup>, que puede ser expuesto de la siguiente forma: ¿cuándo un predicado puede proyectarse válidamente de un conjunto de casos a otros? O ¿cuándo una hipótesis es confirmada por sus instancias positivas? Es necesario distinguir entre predicados naturales y aquellos que se construyen de manera artificial. Si se desarrollara una teoría de la proyección, podría tener lugar en la pragmática de los condicionales, los criterios pragmáticos para medir las propiedades inductivas de los predicados ofrecerían los criterios pragmáticos para ordenar los mundos posibles; (3) Existe una estructura paralela entre el condicional lógico y las funciones de probabilidad condicionales, esto sugiere una conexión entre la lógica inductiva y el condicional lógico. Una asignación de probabilidad y una *s*-función, son formas distintas de describir la relación inductiva entre proposiciones, pero una teoría que enmarque la conexión entre el condicional y la inducción lógica, podría esclarecer las descripciones que da cada una de ellas.

El tercer problema, el problema epistemológico, se basa en la inconformidad que puede tener el empirismo con respecto a que una teoría trate a los contrafácticos como sentencias literales de situaciones no actuales. Se ha dicho que los contrafácticos parecen ser afirmaciones contingentes, y las oraciones contingentes deberían estar fundamentadas en evidencia, pero la evidencia de los contrafácticos no puede ser garantizada sólo por este universo. Para ello, Stalnaker se propone mostrar cómo es que los mundos posibles pueden ser sujetos de una investigación empírica.

---

<sup>34</sup> *Ibid.* P. 110-111.

<sup>35</sup> Nelson Goodman, *The structure of Appearance*, Cambridge, Harvard University Press, 1951.

Podemos definir un mundo posible ignorando algunas verdades de ese mundo, es decir, podemos atribuir características del mundo actual: “Pienso en un mundo posible en el que la población de la Facultad de Filosofía y Letras es de la misma manera que lo es en el mundo actual”, sin embargo, habrá característica de ese mundo que no llegaré a saber si son verdad o no.

Los contrafácticos suelen ser acerca de mundos posibles que son similares al mundo base, lo cual genera información relevante para sus condiciones de verdad. Pensemos en el siguiente ejemplo: Me pregunto acerca de cómo hubieran sido las cosas si en lugar de estudiar filosofía hubiera estudiado veterinaria, el mundo posible es exactamente igual que el mundo actual hasta el punto en el que comienzo a estudiar filosofía, ya que en el mundo posible habría estudiado veterinaria. Dado que tengo noticia acerca de mis disposiciones en el mundo base, eso proporciona un hito de información sobre el mundo posible, lo cual no significa que dicha información sea decisiva, sin embargo, es relevante para saber de la situación no actual que se está analizando.

La teoría semántica de los condicionales nos permite hablar de situaciones posibles no actuales y poder tener conocimiento de ellas a partir de la información proporcionada por el mundo base y su relación con el mundo posible.

El operador  $>$  permite hablar de situaciones posibles que no son actuales/reales y es precisamente de esta forma como Stalnaker entiende los condicionales contrafácticos, como situaciones que no pertenecen al mundo base. Tomando en cuenta esta interpretación de los condicionales contrafácticos, estos coinciden con los condicionales indicativos que hablan de situaciones posibles no actuales, lo que facilita que las condiciones de verdad de ambos condicionales se apliquen de manera similar. Tomemos como ejemplo los siguientes dos condicionales: (1) Si Hilary es presidente de EUA, no construye un muro fronterizo; (2) Si Hilary hubiera sido presidente de EUA, no habría construido un muro fronterizo. Las condiciones para que ambos condicionales sean verdaderos en  $w$  es que, en un mundo seleccionado por la función, en el cual es verdad que Hilary es presidente, también sea verdad que EUA no construye un muro fronterizo; en ambos casos el mundo que se seleccionará será el más similar a  $w$  o será  $w$  mismo en el caso en el que el antecedente es verdadero en  $w$ . De esta manera Stalnaker sostiene que su teoría permite analizar cualquier tipo de estructura condicional.

Al tener las mismas condiciones de verdad, la diferencia entre los condicionales indicati-

vos y los contrafácticos, queda reducida a las intenciones del hablante. El sentido contrafáctico de un condicional sería simplemente la intención del hablante de mostrar su creencia sobre la falsedad del antecedente, lo que implicará que los condicionales se considerarán como indicativos o contrafácticos dependiendo de la creencia que tenga el hablante sobre el antecedente.

A simple vista, parece que la teoría cumple con el objetivo de unificar la evaluación de los condicionales, sin embargo, se pueden formular ejemplos que muestran que la evaluación del condicional indicativo y la de su análogo contrafáctico no siempre coinciden. El ejemplo planteado por Adams es el siguiente<sup>36</sup>: (1) Si Oswald no mató a Kennedy, entonces alguien más lo hizo. (2) Si Oswald no hubiera matado a Kennedy, entonces alguien lo habría hecho. Evaluar estos dos condicionales desde la teoría de Stalnaker nos daría como resultado decir que ambos son verdaderos si el consecuente es verdadero en el mundo seleccionado o que ambos son falsos si el consecuente resulta falso en el mundo seleccionado, el problema que surge es que difícilmente se podría defender que (1) y (2) tienen el mismo valor de verdad, mientras que (1) es indiscutiblemente verdadera (ya que es un hecho que Kennedy está muerto), la mayoría no aceptaría que también (2) es verdadera debido a que entraña la idea de que la muerte de Kennedy tenía que ser necesariamente un asesinato.

La teoría de Stalnaker falla con cualquier par de condicionales que tengan la estructura del ejemplo propuesto por Adams, “Si Jack el destripador no mató a Mary Ann, entonces alguien más lo hizo” y “Si Jack el destripador no hubiera matado a Mary Ann, alguien más lo habría hecho”, “Si Quine no escribió *Los métodos de la lógica*, alguien más lo hizo” y “Si Quine no hubiera escrito *Los métodos de la lógica*, alguien más lo habría hecho”, etc. La característica de estos ejemplos es una idea de necesidad entrañada en el antecedente de la formulación contrafáctica que el condicional indicativo no tiene. Por ello, una teoría adecuada sobre condicionales debe mantener la distinción entre condicionales indicativos y contrafácticos.

Gladys Palau expone en su texto *Condicionales contrafácticos: condiciones de verdad y semántica de mundos posibles de las teorías de R. Stalnaker y D. Lewis*<sup>37</sup>, otra de las fallas que tiene la teoría de Stalnaker<sup>38</sup>, y es que no permite distinguir entre diferentes tipos de condicionales contrafácticos: (1) Si Quine no hubiera escrito *Los métodos de la lógica* entonces

---

<sup>36</sup>Adams, E. "Subjunctive and indicative conditionals." *Formulations of language* 6, p. 89-94, 1970.

<sup>37</sup>Palau, G. Condicionales contrafácticos: condiciones de verdad y semántica de mundos posibles. Acerca de las teorías de R. Stalnaker y D. Lewis..<sup>en</sup> *Critica*, vol.XII No. 34, México. 1980.

<sup>38</sup>*Ibid.* P. 12.

(necesariamente) alguien más lo habría hecho. (2) Si Quine no hubiera escrito *Los métodos de la lógica* entonces (tal vez) alguien más lo habría hecho. Alguien podría rechazar la verdad del primer condicional pensando en que *Los métodos de la lógica* bien podría no haber sido escrito, pero aceptar como verdadero el segundo condicional en el que plantea como posible la idea de que alguien más podría haberlo escrito.

Una tercera dificultad de la teoría se presenta cuando decimos que la función-selección selecciona sólo un mundo. La teoría nos dice que el mundo seleccionado sólo difiere del mundo base en lo que respecta al antecedente, sin embargo, representa una dificultad sostener que es posible que dos mundos varíen únicamente en lo que respecta al antecedente y permanecer igual en el resto de las cosas, ya que con el simple hecho de que en un mundo se cumpla un hecho que es falso en el mundo base hace que ese mundo difiera también de otras formas con el mundo base. Pensemos en una proposición como: “Los humanos nacen de huevos”; esta proposición es falsa en nuestro mundo actual, pero es verdadera en un mundo  $w$ , nuestro mundo actual y el mundo  $w$  difieren en lo que respecta a la proposición mencionada, lo cual implica que en  $w$  la proposición “Los humanos son mamíferos” no podría ser verdadera mientras que sí es verdadera en nuestro mundo actual. Esto nos lleva a aceptar que no es posible que dos mundos varíen en un solo hecho, por lo que existen diversas formas en que un mundo puede variar con respecto al mundo base y no podemos aseverar la unicidad del mundo más similar él<sup>39</sup>.

Esto no sólo representa un problema en las condiciones de evaluación de la teoría de Stalnaker, sino que también lleva a aceptar como verdad lógica el Principio del Tercio Excluido Condicional<sup>40</sup>. Si decimos que el mundo máximamente similar a nuestro mundo base es uno y sólo uno, y que además en ese mundo vale el antecedente, parece prudente que en ese mundo el consecuente tiene que ser verdadero o falso; pero si negamos la unicidad del mundo posible más similar, podemos tener dos mundos que sean similares al mundo base en los cuales se cumpla el antecedente, pero en uno el consecuente es verdadero y en otro es falso. Hasta ahora el supuesto de unicidad no parece representar ningún problema, sin embargo, se pueden formular ejemplos en los que dicho supuesto no parece ser legítimo<sup>41</sup>: Si Bizet y Verdi hubiesen sido compatriotas, Bizet habría sido italiano; Si Bizet y Verdi hubiesen sido compatriotas, Bizet no habría sido italiano. Bajo este ejemplo, parece que no hay nada que nos justifique a decir que el mundo donde

---

<sup>39</sup> *Ibid.* P. 13.

<sup>40</sup> El principio de tercio excluido condicional nos dice lo siguiente:  $(A > B) \vee (A > \neg B)$ .

<sup>41</sup> Gómez Torrente, Mario. “La teoría semántica de las lógicas de condicionales de Stalnaker y Lewis.” En *Agora* 12/2 p. 57-78, 1993. P. 67.

ambos son italianos es más similar al mundo base que el mundo donde ambos son franceses, así como tampoco podríamos justificar lo contrario, ambos mundos tendrían el mismo grado de similitud con el mundo base, lo que haría que se pusiera en duda el principio de unicidad y con él el principio de tercio excluso condicional. Podemos decir que el problema que subyace a la teoría de Stalnaker es el supuesto de unicidad, por ello Lewis construye una propuesta que abandonará este supuesto con la intención de superar las dificultades que representa.

## 2.2. La teoría de David Lewis

La propuesta de David Lewis, expuesta principalmente en su texto *Counterfactuals*<sup>42</sup>, consiste en formular una teoría únicamente del condicional contrafáctico, a diferencia de Stalnaker quien pretendía dar cuenta de una teoría general del condicional, partiendo de la idea de que el condicional indicativo y el condicional contrafáctico son diferentes y no pueden tener el mismo método de evaluación.

Una de las motivaciones que llevó a Lewis a realizar este trabajo fueron las deficiencias que presentaba la implicación estricta al intentar dar cuenta de la semántica de condicionales contrafácticos. Si analizáramos un contrafáctico a partir de la semántica del condicional estricto, significaría que los contrafácticos son verdaderos si y sólo si el condicional material de dicho contrafáctico es verdadero en todos los mundos posibles de cierto tipo. Para Lewis este análisis es incorrecto debido a que tanto el condicional material como la implicación estricta presentan una estructura monotónica<sup>43</sup>, mientras que el condicional contrafáctico parece no atender a dicha estructura; para mostrar cómo es que los contrafácticos no atienden a dicha estructura, pensemos en el siguiente ejemplo: “Si Claudia hubiera asistido a la reunión, la reunión hubiera sido muy divertida, pero si Claudia y Lucía hubieran asistido a la reunión, la reunión hubiera sido muy aburrida, pero si Claudia, Lucía y David hubieran asistido a la reunión, la reunión hubiera sido muy divertida, pero...”<sup>44</sup>. Agregar información al antecedente de un contrafáctico puede afectar el resultado en su consecuente y dar resultados contradictorios entre uno y otro. Teniendo conocimiento de alguna enemistad entre Claudia y Lucía, y algún tipo de carisma entre ambas por David, podríamos intentar asegurar sobre si la reunión habría sido o no divertida.

<sup>42</sup>Lewis, David. *Counterfactuals*. Cambridge: Harvard University Press, 1973.

<sup>43</sup>La implicación estricta, al igual que el condicional material, tiene una estructura monotónica, es decir que si se agregan más fórmulas al conjunto de premisas, el conjunto de sus consecuencias que se pueden deducir no se reducen, lo que la hace inadecuada para el análisis de algunas estructuras condicionales.

<sup>44</sup>Podría representarse de la siguiente forma:  $\varphi_1 \Box \rightarrow \psi$ ,  $\varphi_1 \wedge \varphi_2 \Box \rightarrow \neg \psi$ ,  $\varphi_1 \wedge \varphi_2 \wedge \varphi_3 \Box \rightarrow \psi (\dots)$  Si tuviéramos una estructura monotónica el consecuente se conservaría igual sin importar lo que agreguemos en el antecedente.

Aunque Lewis establece que los condicionales contrafácticos no son condicionales estrictos, sugiere que los contrafácticos son lo que él llama *condicionales variablemente estrictos*. El valor de verdad de un condicional contrafáctico depende del valor que tiene el consecuente en los mundos más parecidos al actual en los que se da el antecedente, y no del valor de verdad del consecuente en los mundos de cierto tipo en los que se da el antecedente (como ocurre con los condicionales estrictos); es por esto que los condicionales estrictos no son suficientes para evaluar los condicionales contrafácticos ya que no delimitan los mundos apropiados para cada uno de estos condicionales.

El sistema de Lewis hace uso del conjunto de mundos posibles, en el cual se incluye el mundo base. De acuerdo con un condicional variable estricto, necesitamos una asignación a cada mundo  $w$  de un conjunto  $\$w$  de esferas de accesibilidad alrededor de  $w$ . Así, el sistema de esferas utilizado para interpretar contrafácticos, es el siguiente<sup>45</sup>:

Sea  $\$$  una asignación para cada mundo posible  $w$  de un conjunto  $\$w$  del conjunto de mundos posibles.  $\$$  es un sistema (centrado) de esferas, y los miembros de cada  $\$w$  son llamados esferas alrededor de  $w$ , si y sólo si, para cada mundo  $w$ , las siguientes condiciones se sostienen:

(C)  $\$w$  está centrado en  $w$ ; esto es que el conjunto  $w$  tiene a  $w$  como el único miembro perteneciente a  $\$w$ .

(1)  $\$w$  es inclusivo; esto es que, cuando  $S$  y  $T$  pertenecen a  $\$w$ , o bien  $S$  está incluido en  $T$  o bien  $T$  está incluido en  $S$ .

(2)  $\$w$  está cerrado bajo unión; es decir, cuando  $S$  es un subconjunto de  $\$w$ , y  $\cup\zeta$  es el conjunto de todos los mundos  $z$  tal que  $z$  pertenece a algún miembro de  $S$ ,  $\cup\zeta$  pertenece a  $\$w$ .

(3)  $\$w$  está cerrado bajo intersección (no vacía); es decir, cuando  $S$  es un subconjunto no vacío de  $\$w$  y  $\cap\zeta$  es el conjunto de todos los mundos  $z$  tal que  $z$  pertenece a todos los miembros de  $S$ ,  $\cap\zeta$  pertenece a  $\$w$ .

Para cada mundo  $w$ , no existe mundo más similar a  $w$  que él mismo, por lo que en (C)

---

<sup>45</sup>Lewis, David. *Counterfactuals*. Cambridge: Harvard University Press, 1973. P. 14.

$w$  debe pertenecer a toda esfera (no vacía) alrededor de  $w$ . (1) nos ofrece la condición de que debe ser posible comparar la similitud de los mundos, si no la tuviéramos significaría que  $\$w$  no es inclusivo, por lo que podríamos tener dos esferas,  $S$  y  $T$ , en  $\$w$  y dos mundos,  $v$  y  $z$ , tales que  $v$  se encuentra en  $S$  y no en  $T$ , y  $z$  se encuentra en  $T$  y no en  $S$ ; dado que  $S$  y  $T$  tienen información comparativa con respecto a la similitud con  $w$ , entonces  $v$  sería más similar a  $w$  que  $z$  y al mismo tiempo  $z$  sería más similar a  $w$  que  $v$ . Las condiciones (2) y (3) se justifican dado que las uniones e intersecciones de los conjuntos de mundos necesitan construir el dominio de similitud, los mundos que se encuentren dentro de la unión serán más similares al mundo base que los que se encuentren fuera de ella, de manera análoga pasa lo mismo con la intersección.

Este sistema de esferas está diseñado para llevar la información acerca de la similitud comparativa entre mundos. Los elementos contenidos en cualquier esfera particular son los mundos accesibles a  $w$  y que además no son diferentes de  $w$  más de cierto límite. Del conjunto de todos los mundos accesibles a  $w$  se establece una jerarquización de esferas, los que sean más similares formarán parte de las esferas interiores, mientras que aquellos que sean menos similares formarán esferas, igualmente jerarquizadas, que contendrán a la esfera  $Sw$ ; cada mundo posible tiene un conjunto de todas las esferas accesibles a dicho mundo al cual nos referiremos como  $\$w$ , donde  $w$  es el mundo base.

Dentro del apartado 1.4, Lewis introduce lo que va a llamar el *supuesto límite*, el cual es la asunción de que a medida que tomemos esferas antecedente-permisivas<sup>46</sup>, que contienen antecedente-mundos<sup>47</sup> cada vez más cercanos a  $w$ , en algún momento encontraremos el límite: la esfera antecedente-permisiva más pequeña, que contiene los antecedente mundos más cercanos. El supuesto nos permite simplificar las condiciones de verdad de los contrafácticos: un contrafáctico será verdadero en  $w$ , si y sólo si el consecuente es verdadero en todo antecedente-mundo que es más cercano a  $w$ . Sin embargo, no hay forma de aseverar la verdad de este supuesto, por lo que cuando no hay una esfera antecedente-permisiva más pequeña, las condiciones de verdad equivalen a lo siguiente: si hay esferas antecedente-permisivas, cuando tomamos una cada vez más pequeña, eventualmente llegaremos a alguna en la que el consecuente se sostiene en todo antecedente-mundo.

En su sistema, Lewis introduce dos operadores para los condicionales contrafácticos, los

---

<sup>46</sup>Llamaremos antecedente-permisiva a toda esfera en la que hay al menos un antecedente-mundo.

<sup>47</sup>Llamaremos antecedente-mundo a los mundos posibles en los que el antecedente  $\alpha$  del condicional  $\alpha \supset \beta$  es verdadero.

cuales se pretende que modelen las diferentes construcciones de condicionales contrafácticos del lenguaje natural. La interpretación de dichos condicionales se dará tomando en cuenta que el valor de verdad de un condicional contrafáctico en un mundo posible depende de los valores de verdad del consecuente en los diversos mundos posibles.

Los dos operadores empleados son los siguientes:  $\Box \rightarrow$  (would), leído como: “Si fuera el caso que  $\alpha$ , entonces sería el caso que  $\beta$ ”; y  $\Diamond \rightarrow$  (might), leído como: “Si fuera el caso que  $\alpha$ , entonces podría ser el caso que  $\beta$ ”; tomando como primitivo al contrafáctico ‘would’. Ambos operadores son interdefinibles de la siguiente manera:

- $\alpha\Box \rightarrow \beta =df. \neg(\alpha\Diamond \rightarrow \neg\beta)$
- $\alpha\Diamond \rightarrow \beta =df. \neg(\alpha\Box \rightarrow \neg\beta)$

Ya que se han establecido los elementos que forman parte del sistema, podemos desarrollar las condiciones de verdad para los operadores. Las condiciones de verdad para el operador ‘would’ contrafáctico son las siguientes:  $\alpha\Box \rightarrow \beta$  es verdadero en  $w$  si y sólo si: i) Ningún  $\alpha$ -mundo pertenece a una esfera  $S$  de  $\$w$ ; o ii) Alguna esfera  $S$  en  $\$w$  contiene al menos un  $\alpha$ -mundo y  $\alpha \supset \beta$  vale en todo mundo perteneciente a  $S$ . Estas condiciones de verdad nos dicen lo siguiente: por un aparte, i) nos presenta el caso de verdad por vacuidad, es decir, ya que no hay ningún mundo posible en la esfera  $\$w$  (que son los mundos accesibles desde  $w$ ) en el que el antecedente sea verdadero, el condicional se vuelve verdadero ya que los condicionales contrafácticos con antecedente imposible son vacuamente verdadero; por otra parte, ii) expresa el caso no vacuo en el cual el condicional contrafáctico es verdadero cuando hay al menos una esfera  $\alpha$ -permisiva tal que todo  $\alpha$ -mundo perteneciente a ella será verdadero el condicional  $\alpha \supset \beta$ .

Las condiciones de verdad para el condicional ‘might’ contrafáctico:  $\alpha\Diamond \rightarrow \beta$  es verdadero en  $w$  si y sólo si: i) hay algún  $\alpha$ -mundo que pertenece a  $S$  en  $\$w$ ; y ii) toda esfera  $S$  en  $\$w$ , que contiene al menos un  $\alpha$ -mundo, hay un mundo en el que  $\alpha \wedge \beta$  es verdadero. Bajo estas condiciones, el condicional ‘might’ parece ser verdadero en la mayoría de los casos.

Estos operadores cuentan con algunas características<sup>48</sup>; (a) si el ‘would’ contrafáctico  $\alpha\Box \rightarrow \beta$  no es vacuamente verdadero, entonces el ‘might’ contrafáctico  $\alpha\Diamond \rightarrow \beta$  también es verdadero; (b) Si  $\alpha\Box \rightarrow \beta$  y su opuesto  $\alpha\Box \rightarrow \neg\beta$  son ambos falsos, entonces  $\alpha\Diamond \rightarrow \beta$  y su opuesto

---

<sup>48</sup>Lewis, David. *Counterfactuals*. Cambridge: Harvard University Press, 1973. P. 21-22.



$\alpha \diamond \rightarrow \neg \beta$  son ambos verdaderos; en este caso  $\beta$  es verdadero en alguno de los  $\alpha$ -mundos más cercanos y  $\neg \beta$  es verdadero en los demás; (c) Cuando  $\alpha \square \rightarrow \beta$  es falso y su opuesto  $\alpha \square \rightarrow \neg \beta$  es verdadero,  $\beta$  no se sostiene en ninguno de los  $\alpha$ -mundos más cercanos y por lo tanto es falso; (d) Cuando  $\alpha$  es inconcebible y por lo tanto  $\alpha \square \rightarrow \beta$  es vacuamente verdadero,  $\alpha \diamond \rightarrow \beta$  es falso.

Para explicar algunas de las características antes mencionadas y con fines didácticos colocaré de manera artificial los operadores ‘would’ y ‘might’ dentro de un cuadro de oposición aristotélico de la siguiente manera:

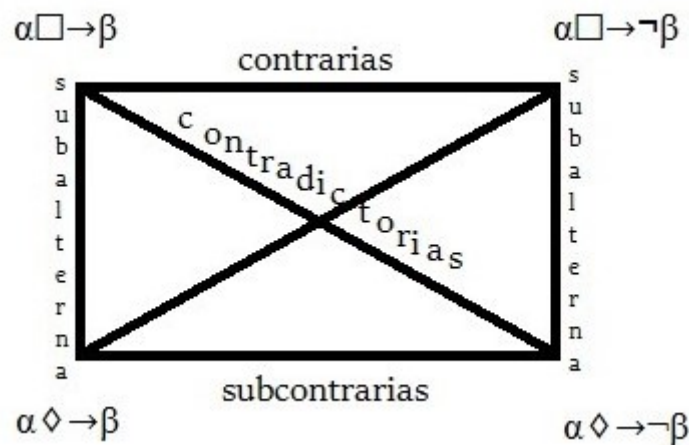


Figura 2: Cuadro aristotélico.

Por una parte, nuestra relación de contradictoriedad nos indica que aquellos juicios o fórmulas que sean contradictorias no pueden tener el mismo valor de verdad, así si  $\alpha \diamond \rightarrow \beta$  es verdadera, entonces  $\alpha \square \rightarrow \neg \beta$  es falsa, o si  $\alpha \square \rightarrow \neg \beta$  es verdadera entonces  $\alpha \diamond \rightarrow \beta$  es falsa, etc. Lo que nos proporciona las definiciones ya mencionadas de los operadores contrafácticos [ $\alpha \square \rightarrow \beta =df. \neg(\alpha \diamond \rightarrow \neg \beta)$  y  $\alpha \diamond \rightarrow \beta =df. \neg(\alpha \square \rightarrow \neg \beta)$ ]. Por otra parte, (a) nos dice que cuando el ‘would’ contrafáctico no sea vacuamente verdadero, el ‘might’ contrafáctico será verdadero, es decir, la relación de subalternas<sup>49</sup> por lo cual es importante la restricción de la verdad por vacuidad, ya que de no ser así no podríamos asegurar la verdad del ‘might’; (b) nos da de nuevo muestra de la relación de contradictoriedad, si los dos ‘would’ son falsos sus contradictorios, los dos ‘might, deberán ser ambos verdaderos; si omitimos de nuevo el caso verdadero por vacuidad del ‘would’, podemos establecer la relación de contrariedad entre  $\alpha \square \rightarrow \beta$  y  $\alpha \square \rightarrow \neg \beta$ , dado que en alguna esfera  $S$  de  $\$w$  tenemos al menos un  $\alpha$ -mundo y  $\alpha \supset \beta$  vale en todo mundo perteneciente a  $S$ , si decimos que  $\alpha \square \rightarrow \beta$  es verdadero, necesariamente su contrario  $\alpha \square \rightarrow \neg \beta$  será falso debido a que en dichas esferas  $S$  que tienen al menos un  $\alpha$ -mundo,  $\neg \beta$  es falso; lo

<sup>49</sup>Dicha relación en el cuadro de oposición aristotélico nos dice que si un juicio universal es verdadero, entonces su juicio particular correspondiente será también verdadero.

anterior también nos da la relación de subcontrariedad, como  $\alpha\Box \rightarrow \beta$  es verdadero,  $\alpha\Diamond \rightarrow \neg\beta$  es falso y su contrario  $\alpha\Diamond \rightarrow \beta$  deberá ser por lo tanto verdadero.

La utilidad de representar los contrafácticos de Lewis en un cuadro de oposición es que nos permite visualizar con mayor facilidad las inferencias posibles que podemos obtener, las cuales están representadas en el cuadro (A) cuando el contrafáctico ‘would’ **no** es vacuamente verdadero y en el cuadro (B) cuando el contrafáctico ‘would’ es vacuamente verdadero:

Supuesto	Inferencias posibles		
$\alpha\Box \rightarrow \beta$ <b>V</b>	$\alpha\Diamond \rightarrow \neg\beta$ <b>F</b>	$\alpha\Diamond \rightarrow \beta$ <b>V</b>	$\alpha\Box \rightarrow \neg\beta$ <b>F</b>
$\alpha\Box \rightarrow \beta$ <b>F</b>	$\alpha\Diamond \rightarrow \neg\beta$ <b>V</b>	-	-
$\alpha\Box \rightarrow \neg\beta$ <b>V</b>	$\alpha\Diamond \rightarrow \beta$ <b>F</b>	$\alpha\Diamond \rightarrow \neg\beta$ <b>V</b>	$\alpha\Box \rightarrow \beta$ <b>F</b>
$\alpha\Box \rightarrow \neg\beta$ <b>F</b>	$\alpha\Diamond \rightarrow \beta$ <b>V</b>	-	-
$\alpha\Diamond \rightarrow \beta$ <b>V</b>	$\alpha\Box \rightarrow \neg\beta$ <b>F</b>	-	-
$\alpha\Diamond \rightarrow \beta$ <b>F</b>	$\alpha\Box \rightarrow \neg\beta$ <b>V</b>	$\alpha\Diamond \rightarrow \neg\beta$ <b>V</b>	$\alpha\Box \rightarrow \beta$ <b>F</b>
$\alpha\Diamond \rightarrow \neg\beta$ <b>V</b>	$\alpha\Box \rightarrow \beta$ <b>F</b>	-	-
$\alpha\Diamond \rightarrow \neg\beta$ <b>F</b>	$\alpha\Box \rightarrow \beta$ <b>V</b>	$\alpha\Diamond \rightarrow \beta$ <b>V</b>	$\alpha\Box \rightarrow \neg\beta$ <b>F</b>

Tabla 6: Tabla (A).

Supuesto	Inferencias posibles
$\alpha\Box \rightarrow \beta$ <b>V</b>	$\alpha\Diamond \rightarrow \neg\beta$ <b>F</b>
$\alpha\Box \rightarrow \beta$ <b>F</b>	$\alpha\Diamond \rightarrow \neg\beta$ <b>V</b>
$\alpha\Box \rightarrow \neg\beta$ <b>V</b>	$\alpha\Diamond \rightarrow \beta$ <b>F</b>
$\alpha\Box \rightarrow \neg\beta$ <b>F</b>	$\alpha\Diamond \rightarrow \beta$ <b>V</b>
$\alpha\Diamond \rightarrow \beta$ <b>V</b>	$\alpha\Box \rightarrow \neg\beta$ <b>F</b>
$\alpha\Diamond \rightarrow \beta$ <b>F</b>	$\alpha\Box \rightarrow \neg\beta$ <b>V</b>
$\alpha\Diamond \rightarrow \neg\beta$ <b>V</b>	$\alpha\Box \rightarrow \beta$ <b>F</b>
$\alpha\Diamond \rightarrow \neg\beta$ <b>F</b>	$\alpha\Box \rightarrow \beta$ <b>V</b>

Tabla 7: Tabla (B).

Existen al menos dos casos que podrían resultar problemáticos: contrafácticos con antecedentes imposibles y, lo que Lewis llama, contrafácticos con antecedentes verdaderos. Una de las condiciones de verdad para el contrafáctico ‘would’ es que puede ser vacuamente verdadero, lo que fácilmente nos permite que un contrafáctico con antecedente imposible sea verdadero. Los contrafácticos con antecedentes imposibles resultan problemáticos porque dan la impresión de poder hacer verdadera a cualquier cosa, cualquier antecedente imposible implicaría lógicamente cualquier consecuente. Estas oraciones podrían no ser considerados como contrafácticos sino condicionales subjuntivos de otro tipo, o podrían ser considerados como contrafácticos no

vacuos, esto último podría justificarse de la siguiente manera: así como hay *posibles* mundos posibles que se diferencian del mundo base por hechos contingentes o empíricos, del mismo modo podríamos suponer que existen *imposibles* mundos posibles que difieren del mundo base en las verdades lógicas, matemáticas y filosóficas. Este análisis explica cómo podemos diferenciar el valor de verdad entre diferentes contrafácticos con antecedentes imposibles.

Aunque esta última propuesta parece resolver los problemas que vienen de tener contrafácticos con antecedentes imposibles, Lewis no piensa que sea necesario diferenciar el valor de verdad de los distintos condicionales con antecedentes imposibles, sin embargo, es verdad que hay algunos que estamos dispuestos a aceptar y otros que no. Para ilustrar lo anterior Lewis enuncia<sup>50</sup> los siguientes dos pares de contrafácticos<sup>51</sup>:

*Si existiera el mayor número primo  $p$ ,  $p!+1$  sería también un primo.*

*Si existiera el mayor número primo  $p$ ,  $p!+1$  sería compuesto.*

*Si existiera el mayor número primo  $p$ , habría seis sólidos regulares.*

*Si existiera el mayor número primo  $p$ , los cerdos tendrían alas.*

Mientras que el primer par de contrafácticos parecen sensatos, el segundo par de contrafácticos no serían tan fácilmente aceptados como verdaderos. El aceptar una postura de este tipo nos obligaría a dar una explicación sobre por qué es que ciertos contrafácticos con antecedentes imposibles los tomamos por verdaderos mientras que otros con la misma estructura no los consideramos verdaderos. Del mismo modo tenemos contrafácticos con antecedente verdadero que no estamos dispuestos a aceptar, pero tampoco su negación. Por estas razones Lewis no tiene problema en hacer vacuamente verdaderos a los contrafácticos con antecedentes imposibles aunque es consciente de la poca aceptación que esto podría tener por lo que define un ‘would’ más fuerte que no pueda ser vacuamente verdadero y que tiene las siguientes condiciones de verdad:

$\alpha \Box \Rightarrow \beta$  es verdadero en el mundo  $w$  (de acuerdo al sistema de esferas  $\$$ ) si y sólo si hay alguna esfera  $S$  en  $\$w$  tal que  $S$  contenga al menos un  $\alpha$ -mundo, y  $\alpha \supset \beta$  se sostiene en todos los mundos de  $S$ .

---

<sup>50</sup> *Op cit.* P. 25.

<sup>51</sup> De acuerdo a cómo hemos definido los contrafácticos dentro de este trabajo, pareciera que los cuatro condicionales enunciados no son contrafácticos. Esto se debe a distinciones que tenemos dentro en español y que parecen no darse en inglés, recordamos que el texto de Lewis fue escrito en inglés.

Debido a la interdefinibilidad del ‘would’ y el ‘might’, se introduce un ‘might’ débil que es vacuamente verdadero cuando el antecedente es imposible y se define de la siguiente forma:  $\alpha \diamond \Rightarrow \beta = \text{df. } \neg(\alpha \Box \Rightarrow \neg\beta)$ , y tiene las siguientes condiciones de verdad:

*$\alpha \diamond \Rightarrow \beta$  es verdadero en el mundo  $w$  (de acuerdo a un sistema de esferas  $\$$ ) si y sólo si toda esfera  $S$  en  $\$w$  que contiene al menos un  $\alpha$ -mundo, contiene al menos un  $\alpha$ -mundo en el que  $\alpha \wedge \beta$  se sostiene.*

Estas condiciones de verdad para el ‘might’ débil nos dan como resultado las mismas cuestiones que nos motivaron en un principio a proponer el ‘would’ más fuerte. Por eso Lewis no tiene problema en dejar los contrafácticos con antecedentes imposibles como vacuamente verdaderos.

Por otra parte, tenemos los contrafácticos con antecedentes verdaderos, que parecen representar un defecto para la teoría. Sin embargo, para Lewis no necesariamente es un defecto que provoque que sean automáticamente falsos o ambiguos, por la simple razón de que alguien podría enunciar legítimamente un contrafáctico con antecedente verdadero debido a que desconoce la verdad del antecedente y lo cree falso. Por ello no tiene problemas en plantear que los contrafácticos con antecedentes verdaderos sean reducidos a condicionales materiales.

Pensemos que el antecedente  $\alpha$  es verdadero en el mundo  $w$ , entonces hay una esfera permisiva alrededor de  $w$ , ya que  $w$  es una esfera. Cuando el consecuente  $\beta$  es verdadero en  $w$ , entonces hay una esfera  $\alpha$ -permisiva en la cual  $\alpha \supset \beta$  se sostiene; por lo que  $\alpha \Box \rightarrow \beta$  es verdadero en  $w$ ; de igual forma, cada esfera  $\alpha$ -permisiva alrededor de  $w$  contiene un mundo en el que  $\alpha \wedge \beta$  se sostiene —excepto el conjunto vacío que no contiene esferas  $\alpha$ -permisivas— y contiene también al mismo mundo  $w$ , por lo que  $\alpha \diamond \rightarrow \beta$  es verdadero. Por otra parte, si el consecuente  $\beta$  es falso en  $w$  y no hay una esfera  $\alpha$ -permisiva alrededor de  $w$  donde  $\alpha \supset \beta$  se sostenga, entonces  $\alpha \Box \rightarrow \beta$  sería falso; y como hay una esfera  $\alpha$ -permisiva que no contiene a algún mundo en el que  $\alpha \wedge \beta$  se sostenga, entonces  $\alpha \diamond \rightarrow \beta$  es falso.

De igual forma que en la teoría de Stalnaker, los condicionales variablemente estrictos no permiten algunas inferencias que son posibles con el condicional material y con cualquier condicional estricto constante. Las inferencias que quedan invalidadas en los condicionales variablemente estrictos son el refuerzo del antecedente, la transitividad y la contraposición.

Esta teoría, tal como Lewis la presenta, tiene diversas virtudes. Por una parte, se enfoca únicamente a las condiciones de verdad de los contrafácticos y además hace distinción entre el contrafáctico ‘would’ y el ‘might’, cuya interdefinición parece apropiada y es difícil encontrar un contraejemplo que fuese capaz de demostrar la inadecuación de estos en el lenguaje natural. Por otra parte, aunque el contrafáctico ‘might’ podría parecer demasiado permisivo, sus análogos del lenguaje natural también lo son, lo que resulta ser una virtud de la teoría que se adecue al lenguaje natural.

Aunque la teoría parece acertar en muchos puntos que la teoría de Stalnaker no pudo, la propuesta de Lewis no está exenta de dificultades. A pesar de los intentos de Lewis por demostrar lo contrario, el hecho de que los contrafácticos con antecedentes imposibles sean vacuamente verdaderos representa una de ellas. El mismo Lewis acepta que existen contrafácticos con antecedentes imposibles que tiene sentido afirmar y otros que no, pero que esto se debe únicamente a razones conversacionales.

Es importante señalar que estos resultados pueden ser obtenidos si se entiende “imposibilidad” como “imposibilidad lógica”<sup>52</sup> y además se atiende a la noción de deducibilidad en lógica clásica. Si, por otro lado, se entiende “imposibilidad” como algo más distinto de lo lógicamente imposible, entonces las dificultades surgen. Tenemos por ejemplo el caso de la imposibilidad física, que sin lugar a dudas forman parte de los condicionales contrafácticos. Para esclarecer esto tomemos en cuenta algunos ejemplos que presenta Gladys Palau en su texto<sup>53</sup>: (1) Si Urano y Neptuno no hubieran estado sujetos a la gravedad, Leverrier habría descubierto 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. Ambos contrafácticos cuentan con antecedentes físicamente imposibles, al negar aquello que la ley física afirma, pero mientras que (1) es falso, (2) es verdadero. Podemos asegurar la falsedad de (1) porque sabemos que el descubrimiento de Neptuno fue dado a partir de irregularidades en la órbita de Urano las cuales se explicaban por la hipótesis de un planeta exterior y la ley de la gravedad; mientras que (2) es sin lugar a dudas verdadero, ya que Kepler encontró la diferencia entre sus cálculos y las observaciones precisamente porque la órbita de Marte no es circular, ya que si hubiera sido circular la diferencia no existiría. Sin embargo, la propuesta de

---

<sup>52</sup>Con imposibilidad lógica me refiero a aquello que va en contra de alguno de los tres principios de la lógica clásica: identidad, tercio excluido, no contradicción.

<sup>53</sup>Palau, G. Condicionales contrafácticos: condiciones de verdad y semántica de mundos posibles. Acerca de las teorías de R. Stalnaker y D. Lewis..<sup>en</sup> *Critica*, vol.XII No. 34, México. 1980. P. 19.

Lewis asignaría verdadero a ambos condicionales ya que, según la teoría, un contrafáctico de la forma  $\alpha \Box \rightarrow \beta$  es vacuamente verdadero en un mundo  $w$  no sólo cuando es lógicamente imposible, también cuando es lógicamente posible, pero  $w$  no tiene acceso a ningún mundo en el que se dé esa posibilidad. Lewis tenía en cuenta que los contrafácticos con antecedente imposible podían resultar problemáticos<sup>54</sup>, por lo que propone otras posibles condiciones de verdad para el operador “would” cuando no fuera verdadero por simple vacuidad, pero por el momento no hablaremos de dichas posibilidades.

Otra dificultad que presenta la teoría es la expuesta por Frank Jackson en su texto “A causal theory of counterfactuals”<sup>55</sup>. Sostiene que las condiciones de verdad en la teoría de Lewis no son apropiadas. Jackson se encarga de exponer cómo es que las condiciones de verdad de Lewis hacen verdaderos algunos condicionales contrafácticos que son indudablemente falsos. Los ejemplos presentados por Jackson como contraejemplos a la teoría tienen la forma de los ejemplos tratados en el párrafo anterior que hablan de imposibilidad física, sin embargo, su objeción no se reduce únicamente a estos ejemplos. Imaginemos un mundo no determinista, es decir que todas las cosas que ocurren en él no tuvieran causas ni efectos, sino que fueran completamente accidentales, en este mundo ningún *contrafáctico estándar*<sup>56</sup> sería verdadero y todos los contrafácticos “might” serían verdaderos. La teoría de Lewis tomaría esto como falso, ya que podríamos seleccionar los mundos accesibles más similares a ese mundo no determinista y a partir de eso analizar el valor de verdad del contrafáctico en cuestión.

Una cuestión tal vez no tan relevante es preguntar por las razones que llevaron a Lewis a no añadir los subjuntivos futuros a la teoría. A simple vista parece que extender la teoría a los subjuntivos futuros no genera ninguna dificultad, a un condicional que tenga como antecedente un subjuntivo futuro falso en un mundo  $w$ , determinaríamos su valor de verdad eligiendo una esfera de mundos posibles y elegir la que tenga mayor similitud con  $w$  en el que el antecedente sea verdadero y así determinar qué ocurre con el condicional (tal y como ocurre con los subjuntivos no futuros).

Así tenemos que el sistema de Lewis adquiere mayor expresividad con respecto al sistema de Stalnaker, además de que su sistema de esferas nos da una mejor forma de evaluación,

---

<sup>54</sup>Sin embargo, no podemos asegurar que las dificultades que estuviera considerando sean las mismas que consideramos en los ejemplos anteriores.

<sup>55</sup>En *Australasian Journal of Philosophy*, vol. 55, No. 1, 1977.

<sup>56</sup>Son aquellos que cumplen las siguientes condiciones: 1)  $\neg P$ ; 2)  $\neg \Box(P \supset Q)$ ; 3)  $P$  y  $Q$  no son contrafácticos.

sin embargo, su teoría no está exenta de dificultades, las cuales pueden poner en duda si es la mejor forma de evaluar los contrafácticos.

## 2.3. Análisis comparativo

Tanto la teoría de Robert Stalnaker como la de David Lewis están basadas en las semánticas de mundos posibles y la similitud entre ellos, por ello se suelen estudiar de manera conjunta. A pesar de que ambas teorías puedan seguir en algunos aspectos la misma línea argumentativa, también existen claras diferencias entre ambas. Aunque estas diferencias ya se han dejado entrever en el desarrollo de los apartados anteriores, los dos siguientes apartados estarán dedicados a desarrollar puntualmente las similitudes y las diferencias que existen en estas teorías.

### 2.3.1. Similitudes entre las teorías

Además de que ambas teorías se apoyan en la semántica de mundos posibles, como se mencionó anteriormente, también rechazan inferencias que son válidas para el condicional material y el condicional estricto, a saber (i) el refuerzo del antecedente, (ii) la transitividad y (iii) la contrapositiva.

Por una parte, (i) queda invalidado debido a que agregar información al antecedente de un contrafáctico puede alterar el resultado del consecuente: a) “Si hubiera llevado paraguas, no me habría mojado”; b) “Si hubiera llevado paraguas y lo hubiera olvidado en la oficina, me habría mojado”. Al reforzar el antecedente de a) nos da como resultado el contrafáctico b) que puede ser igualmente verdadero pero tiene como consecuente la proposición contraria al consecuente del contrafáctico a).

La invalidez de (i) tiene como consecuencia que (ii) sea también una inferencia inválida, dado que el refuerzo del antecedente se sigue de la transitividad<sup>57</sup>. Otra de las razones por las cuales (ii) no es una inferencia de ambos sistemas es debido a las condiciones de verdad. En la teoría de Stalnaker la función-selección puede seleccionar diferentes mundos para cada uno de los diferentes antecedentes, mientras que en la teoría de Lewis las esferas de accesibilidad pueden ser distintas para cada uno de los antecedentes seleccionados dando así consecuentes distintos.

---

<sup>57</sup>Supongamos que  $A \Box \rightarrow B$  y sea  $(A \wedge C) \Box \rightarrow A$  (por ley lógica), por transitividad podríamos decir que  $(A \wedge C) \Box \rightarrow B$ , que resulta en el refuerzo del antecedente de nuestra primera proposición.

El rechazo de (iii) tiene como explicación lo que mencionamos anteriormente en (ii), diferentes antecedentes pueden dar como resultado diferentes mundos posibles en los que se va a evaluar el contrafáctico. Pensemos en el siguiente ejemplo: a) Si Héctor no hubiese estudiado diseño gráfico, no viviría en la CDMX; b) Si Héctor hubiera vivido en la CDMX, habría estudiado diseño gráfico. La verdad de a) no nos permite inferir necesariamente la verdad de b), podríamos tener que a) es verdadero de acuerdo al mundo posible más cercano al que se tiene acceso, o a la esfera de accesibilidad, pero eso no significa que al evaluar b) vayamos a tener como resultado aquellos mundos en los que Héctor estudia diseño gráfico.

### 2.3.2. Diferencias entre las teorías

Varias de las diferencias que podemos enunciar que existen entre las teorías son expuestas por el mismo Lewis, ya que una de las intenciones de éste es dar un mejor análisis a los contrafácticos del que le da Stalnaker.

1. La teoría de Lewis distingue entre los condicionales ‘would’ y ‘might’ mientras que la teoría de Stalnaker no es capaz de distinguirlos, lo que hace que la teoría de Lewis tenga mayor expresividad.
2. En la teoría de Stalnaker existe solamente un mundo, en el que se da el antecedente, que es el más cercano al mundo base, mientras que Lewis utiliza un sistema de esferas y estipula el supuesto límite, el cual le permite tener una esfera antecedente-permisiva que es la más cercana al mundo base. Esto hace que la teoría de Stalnaker se encuentre sobre un supuesto más fuerte que el que tiene la teoría de Lewis.
3. Stalnaker estipula un mundo absurdo ( $\lambda$ ) para poder establecer la vacuidad de los contrafácticos con antecedentes imposibles, mientras que en la teoría de Lewis ese supuesto es innecesario ya que la vacuidad ya se encuentra establecida dentro de las condiciones de verdad del contrafáctico.
4. Recordemos que la teoría de Stalnaker está pensada para establecer condiciones de verdad para los condicionales en general, los condicionales con antecedente verdadero se reducen a condicionales materiales pero aquellos condicionales con antecedente falso no resultan necesariamente verdaderos, como pasa en el condicional material, sino que pueden tener valor de verdad verdadero o falso, es decir, que al evaluar un condicional con antecedente falso, se selecciona un mundo donde el antecedente es verdadero y puede darse el caso en que el consecuente resulte falso, el condicional será falso en el mundo base a pesar de que su



antecedente sea falso<sup>58</sup>. La teoría de Lewis está enfocada únicamente en los condicionales contrafácticos, y aunque en su teoría los contrafácticos con antecedentes verdaderos se reducen a condicionales materiales, no significa que el operador contrafáctico se defina a partir del condicional material.

5. La teoría de Stalnaker acepta el principio del tercio excluso contrafáctico:  $(\alpha \Box \rightarrow \beta) \vee (\alpha \Box \rightarrow \neg\beta)$ , el cual parece sensato, en el caso vacuo ambos disyuntos son verdaderos; en el caso no vacuo el mundo que selecciona la función donde el antecedente es verdadero es sólo uno y en él se da  $\beta$  o bien se da  $\neg\beta$ . Sin embargo, en la teoría de Lewis este principio no es válido ya que es posible que ambos disyuntos sean falsos. La teoría de Lewis no selecciona un único mundo posible más cercano al mundo base, sino que selecciona una esfera de accesibilidad que puede contener diversos mundos posibles todos con el mismo grado de similitud al mundo base, y dentro de esa esfera puede haber un  $\alpha$ -mundo en el que  $\beta$  sea falso y otro  $\alpha$ - mundo en el que  $\neg\beta$  sea falso, lo que haría que tanto  $\alpha \Box \rightarrow \beta$  como  $\alpha \Box \rightarrow \neg\beta$  sean ambos falsos.

## 2.4. Conclusiones

Hemos presentado estas dos teorías que representan parte importante del análisis de los condicionales contrafácticos, la teoría de Stalnaker que, a pesar de pretender ser una teoría para los condicionales en general, planta bases para el estudio de los condicionales contrafácticos, y la teoría de Lewis que presenta un estudio más fino y elaborado sobre los contrafácticos.

A pesar de los esfuerzos de ambos autores, las estructuras de los condicionales contrafácticos pueden ser muy variables, por lo que pareciera que este análisis no resulta del todo satisfactorio en algunos tipos de contrafácticos como los presentados por Jackson (mencionados en el apartado de Lewis como objeciones a la teoría), los cuales parecen atender a la idea de causalidad que no se refleja en la similitud de mundos posibles.

Estas cuestiones dejan abierta la pregunta de si es posible dar una teoría uniforme a los contrafácticos o si es necesario hacer una distinción de los tipos de contrafácticos y así asignarle a cada uno de ellos un análisis distinto de sus formas de evaluación.

---

<sup>58</sup>Esto tiene relevancia cuando evaluamos en la teoría de Stalnaker condicionales indicativos con antecedente falso, ya que podrían resultar falsos independientemente de la falsedad de su antecedente.

### 3. La teoría de Eric Hiddleston

Eric Hiddleston desarrolla<sup>59</sup> una propuesta para la evaluación de condicionales contrafácticos basada en “modelos causales”, la cual difiere de la teoría Stalnaker-Lewis. El rechazo a dicha teoría se debe a que la forma en que es presentada la “similitud” de mundos posibles no hace referencia a las relaciones causales, necesidades naturales, etc., las cuales, para Hiddleston, son cruciales en la evaluación de contrafácticos.

De manera general, para evaluar un condicional contrafáctico de la forma “Si A hubiera sido, entonces B habría sido”, se necesita empezar con una red causal de eventos en los que de hecho A no es el caso, por lo que se necesita realizar un “rompimiento mínimo” en la red causal que va a permitir que A sea verdadero. Posteriormente, se trazan las consecuencias causales del rompimiento y se mantienen aquellas que no fueron influenciadas por él; “Si A hubiera sido, entonces B habría sido” es verdadero si y sólo si B es verdadero en dichas situaciones.

Bajo esta idea, Hiddleston presenta diversos ejemplos que sirven como contraejemplos para las teorías basadas en la similitud de mundos posibles, para después dar pie al desarrollo de su teoría, la cual pretende abarcar un mayor número de casos que la teoría Stalnaker-Lewis, además de arrojar resultados más satisfactorios.

#### 3.1. Objeciones a la teoría Stalnaker-Lewis

La forma en que Hiddleston presenta las condiciones de verdad para los contrafácticos dadas por Lewis es la siguiente: “ $A > B$  es verdadero si y sólo si B es verdadero en los mundos en los que A es verdadero y que son más similares al actual”. Su objeción radica en que Lewis no especifica a qué es lo que se refiere con “similitud” y en lugar de eso sólo presenta la “resolución estándar de vaguedad” para cualquier similitud y la presenta de la siguiente manera<sup>60</sup>:

1. Es de primera importancia evitar violaciones de leyes generalizadas, diversas y grandes.
2. Es de segunda importancia maximizar la región espacio temporal a través de la cual prevalece la perfecta combinación de hechos particulares.
3. Es de tercera importancia evitar incluso las pequeñas, localizadas y simples violaciones de la ley.

---

<sup>59</sup>Hiddleston, Eric. “A causal theory of counterfactuals”. *Noûs*, Vol. 39, No. 4 (Dec., 2005), p. 632-657.

<sup>60</sup>*Ibid.* P.633.

4. Es de poca o ninguna importancia asegurar una similitud aproximada de un hecho particular, incluso en asuntos que nos preocupan mucho.

Esto tiene como consecuencia la idea de que al evaluar un contrafáctico de la forma  $A > B$  podemos mantener fijo el pasado de  $A$  teniendo solamente algunas divergencias menores que permiten a  $A$  ser verdadero, (cláusulas (2) y (3)); se supone a  $A$  como verdadera y se deja que las cosas ocurran como las leyes dictan, (cláusula (1)).

Pasaremos a ver los ejemplos<sup>61</sup> proporcionados por Hiddleston en los cuales la propuesta de Lewis pareciera ofrecernos una evaluación equivocada. El primer tipo de ejemplos es aquel que involucra antecedentes que son causalmente irrelevantes con sus consecuentes:

*Ejemplo 1:* Alice le ofrece a Ben una apuesta en un lanzamiento de moneda, justo e indeterminado. Ben apuesta por cruz. Alice arroja la moneda, y aterriza cara. Ben pierde. Alice le dice a Ben. "Si hubieses apostado cara, habrías ganado". Lo cual parece verdadero.

*Ejemplo 2:* Hay una caja negra con una fuente que produce un fotón cada 10 segundos. Cada fotón viaja hacia un trozo de vidrio en la caja, y los detectores registran si se transmite (probabilidad 0.8) o si se refleja (probabilidad 0.2). Las luces en la caja registran los resultados: una luz verde para transmisión, una luz roja para reflexión. Alice y Ben tienen una política de que Ben siempre apuesta \$1 a la reflexión (cuando apuesta), con una probabilidad justa de 4:1. Ben toma un descanso del juego y va a tomar un refresco de la máquina al final del pasillo. Mientras él no está, la luz roja (improbablemente) enciende 5 veces consecutivas. Cuando Ben regresa, Alice le informa de esta improbable secuencia y le dice, "Si no te hubieras ido a conseguir el refresco, ahora serías \$20 más rico". Lo que parece —de nuevo— verdadero.

Antes de pasar al análisis, Hiddleston nos pide que de momento la cláusula (4) no sea tomada en cuenta, ya que tiene la suposición de que ésta fue agregada por Lewis más tarde debido a los problemas que mencionaremos a continuación. Por una parte, en el primer ejemplo, la cláusula (1) nos indica que los mundos similares en lo que Ben apuesta cara deben tener las mismas leyes que el actual, dado que el lanzamiento de la moneda no está determinado por sus causas anteriores; la cláusula (2) nos pide que dichos mundos coincidan exactamente con la historia actual hasta un momento antes de que Ben apueste; y en conjunción con la cláusula

---

<sup>61</sup>Ibíd. P. 634 – 635.

(3), la menor cantidad de eventos (en el pensamiento de Ben) deben ocurrir de manera diferente para que elija cara en lugar de cruz. Como todos los mundos similares se comportan con las leyes del actual y sus aspectos relevantes coinciden, en algunos de esos mundos la moneda cae cruz, lo que hace que la teoría sobre la similitud de mundos haga al contrafáctico del ejemplo 1 falso, pero anteriormente se señaló como verdadero.

El segundo tipo de ejemplos tienen que ver con que sus antecedentes mantienen algún tipo de relación causal de forma positiva con sus consecuentes:

*Ejemplo 3:* Jane es mordida por una serpiente venenosa en una excursión. Hay un antídoto para el veneno de serpiente, pero Jane no tiene dicho antídoto a la mano. Improbablemente, el cuerpo de Jane lucha contra el veneno, y ella sobrevive. El antídoto, aunque bueno, no es perfecto: queda alguna posibilidad de muerte entre las personas que se exponen a ambos, el veneno y el antídoto. Parece verdad que: “Si Jane hubiera tomado el antídoto, entonces ella habría vivido”.

Una vez más la teoría de Lewis hace falso el contrafáctico enunciado en el ejemplo 3. La situación mencionada hace posible que el antídoto no sea efectivo (dado que no es perfecto), así que Jane no sobrevive en todos los mundos posibles y por lo tanto el contrafáctico es falso.

El añadir la cláusula (4) puede proporcionarnos un resultado distinto, podría permitirnos tomar los mundos en los que no se cumplen los consecuentes como más alejados de aquellos en los que sí se cumplen, obteniendo un resultado satisfactorio. Sin embargo, agregar la cláusula (4) también puede dar resultados no satisfactorios a ciertas evaluaciones. Para esto Hiddleston nos proporciona el siguiente ejemplo:

*Ejemplo 4:* Alice le ofrece a Ben una apuesta en lanzar una moneda, pero esta vez Alice puede influir el resultado. Ben apuesta cruz. Esperando ganar, Alice lanza la moneda teniendo una gran probabilidad (0.8, dice) de que caiga cara. El resultado es cara. Ella le dice a Ben, “Si hubieras elegido cara, habrías ganado”. Lo cual parece falso.

Lo que la cláusula (4) nos ofrece es que, por “similitud aproximada”, debemos aplicar el resultado del lanzamiento de la moneda, el cual fue cara, y por lo tanto el contrafáctico en el ejemplo 4 sería verdadero, sin embargo, el modo en el que el ejemplo es presentado nos haría

rechazar dicho resultado; Alice es capaz de influir en el resultado, así que, si Ben hubiera elegido cara, ella hubiera influido de tal manera en que el resultado habría sido cruz.

El siguiente ejemplo, que sería parte grupo del primer tipo, nos proporciona un caso en el que la cláusula (4) no es suficiente para dar un resultado satisfactorio:

*Ejemplo 5:* Es 1973. Nixon se sienta con su dedo sobre el botón de destrucción, observando dentro de un circuito cerrado de televisión cómo Alice y Ben apuestan en lanzar la moneda. Él les dice a sus asesores que presionará el botón si Ben gana la siguiente apuesta, y que de otra manera lo quitará. Ben apuesta cruz, la moneda cae cara, y Ben pierde la apuesta. De mala gana Nixon pone el botón de vuelta en el cajón. Uno de sus asesores suelta un suspiro y piensa para sí mismo, “Si Ben hubiera elegido cara, habría ganado y Nixon nos habría destruido a todos”. Lo que parece verdad.

La “similitud aproximada” nos pide que tomemos en cuenta la semejanza de los mundos con respecto a que ocurra o no un holocausto nuclear, y aún más a la semejanza del lanzamiento de la moneda; sin embargo, Hiddleston nos dice que ninguna de estas situaciones es eficaz dentro de este ejemplo, el cual es similar al ejemplo 1 (dado que son ejemplos del mismo tipo), así que la similitud aproximada tampoco va a ser capaz de explicarlo.

Estos casos han sido motivo de discusión en el tema de los contrafácticos, y mientras que Lewis no les toma importancia y los deja al margen, para Hiddleston son muy importantes ya que proporcionan la base para un enfoque distinto de los contrafácticos.

### **3.2. Desarrollo de la teoría de Hiddleston**

Para Hiddleston hay solamente dos intentos serios por explicar los casos mencionados en el apartado anterior, el de Igal Kvar<sup>62</sup> y el de Stephen Barker<sup>63</sup>. De entre los ejemplos, podemos ver una “obvia” diferencia entre los hechos causales del ejemplo 1 y del ejemplo 4, la cual se refleja en la diferencia de sus relaciones probabilísticas; dentro del ejemplo 1, la apuesta de Ben es probabilísticamente irrelevante para el resultado, mientras que los hechos en 4 no lo son. El intento de Kvar involucra casos que hacen uso de las relaciones probabilísticas. Analicemos

---

<sup>62</sup>Kvar, Igal. *A Theory of Counterfactuals*. Indianapolis: Hackett. 1986.

<sup>63</sup>Baker, Stephen. “Counterfactuals, probabilistic counterfactuals, and causation”. *Mind* 108, 1999.

esto con los siguientes ejemplos<sup>64</sup>

*Ejemplo 6:* Alice le ofrece a Ben una apuesta en lanzamiento de moneda, Ben elige cruz, la moneda cae cara, y Ben pierde la apuesta. (6a): “Si Ben hubiera elegido cara, tomado la moneda de Alice y la tirara él mismo, habría ganado”. (6b): “Si Ben hubiera elegido cara, tomado la moneda en el aire y la tirara él mismo, habría ganado”. Tanto (6a) como (6b) parecen falsos.

Tal como este ejemplo se presenta, no tenemos razones para decir que el hecho de que Ben tome la moneda afecta la probabilidad de que el resultado sea cara, porque no es así, no obstante, establece una inferencia con el resultado de que sea cara. La forma en que Kwart<sup>65</sup> aborda el problema es ofreciendo una teoría probabilística compleja de “irrelevancia causal” y “relevancia causal puramente positiva”. Lo que Hiddleston retoma de esto es la necesidad de apelar a las causas actuales. Dentro del ejemplo 6, las causas actuales serían que: Alice sea quien arroje la moneda, que sea arrojada con cierta fuerza, que recorriera cierta trayectoria, etc. Estas causas constituyen una inferencia con el resultado, ya sea que el resultado de la probabilidad se altere o no.

Otra forma de explicar el problema sería haciendo uso de la noción de “socavar” dada por John Pollock<sup>66</sup> “Decir que P socava Q es decir que P implica la falsedad de toda historia causal de Q”. La idea es que los P-mundos mínimamente alterados van a ser aquellos que se superpongan al máximo con el mundo actual, respecto a los hechos que no-socavados por P. Sin embargo, Barker nos ofrece una explicación más simple y natural:

*Ejemplo 7:* Jane considera apostar \$1000 que los ojos de serpiente aparecerán en un rollo de dos dados de seis caras. Ella se abstiene de apostar, los dados se tiran, y caen ojos de serpiente. Sin que Jane lo sepa, hay electroimanes debajo de la mesa que si son activados evitan que caigan ojos de serpiente. Los imanes son controlados por Mandrake, quién tiene la política de interferir con apuestas de alto riesgo el 5% del tiempo. Considera: “Si Jane hubiera apostado ojos de serpiente [B] y Mandrake no hubiera activado los imanes [ $\neg M$ ], entonces los dados habrían caído ojos de serpiente [S]”,  $(B \wedge \neg M) > S$ . Que parece verdadero.

<sup>64</sup>Hiddleston, Eric. “A causal theory of counterfactuals”. *Noûs*, Vol. 39, No. 4 (Dec., 2005), p. 636-637.

<sup>65</sup>Kwart comienza a desarrollar esta teoría en: Kwart, Igal. *A Theory of Counterfactuals*. Indianapolis: Hackett. 1986, pero continúa desarrollándola en: Kwart, Igal. “Transitivity and Preemption of Causal Relevance”. *Philosophical Studies* 64, p. 125-60, 1991. Y en: Kwart, Igal, “Causal Independence”. *Philosophy of Science* 61, 96, 1994.

<sup>66</sup>Pollock, John. “A refined theory of counterfactuals”. *The journal of philosophical logic* 10, 1981.

Tal como el ejemplo se presenta, que Jane no apostara ( $\neg B$ ) es una causa actual que el resultado de los dados fuera ojos de serpiente (S), y que Jane no apostara es la causa por la que Mandrake no activó los imanes, que a su vez es causa de que el resultado fueran ojos de serpiente. La noción de Pollock nos daría que, B y todo lo que  $(B \wedge \neg M)$  implique, socava a S, haciendo falso el contrafáctico del ejemplo 7 cuando se había dicho verdadero.

Hiddleston intenta solucionar estos problemas agregando la idea de “ruptura causal” del curso de eventos actuales, la cual ocurre en una situación hipotética cuando las causas inmediatas de A ocurren, pero A no. Las diferencias que se pueden presentar desde una realidad “mínimamente alterada” nos permiten rastrear las situaciones de P hasta un rompimiento inicial.

Esto es lo que comienza a dar forma a la teoría que Hiddleston quiere presentar y que se dirigirá a los “modelos causales”<sup>67</sup>. Lo que ofrece es considerar a los contrafácticos usando versiones desmontadas de los “modelos causales”, los cuales pueden ser representados en gráficas<sup>68</sup>, a las que llama “gráficas acíclicas dirigidas”, GADs. Las flechas en las gráficas representan las relaciones causales. Los *padres* de un nodo X,  $pa(X)$ , son los nodos que tienen flechas en ellos. Veamos la gráfica del ejemplo 3: Fig. 1,  $pa(\text{Sobrevivir}) = \{\text{Veneno}, \text{Antídoto}\}$ .

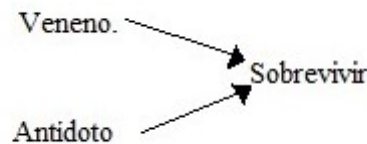


Figura 3: Gráfica acíclica dirigida: Veneno-Antídoto

Los nodos de las gráficas representan las *variables* de los eventos. Los modelos causales incluyen valores de las variables: Veneno = 0 (no veneno) o Veneno = 1 (veneno). De este modo podemos describir las relaciones entre variables por medio de leyes causales: Si Veneno = 1 y Antídoto = 0, entonces  $p(\text{Sobrevivir} = 1) = 0.5$ .

Hiddleston utilizará modelos causales que permitan leyes indeterminadas, estos tendrán

<sup>67</sup>Hiddleston usa como referencia a las teorías presentadas por Judea Pearl en: "Probabilities of Causation: Three Counterfactual Interpretations and Their Identification". Synthese 121, p. 93-149, 1999; y Causality. Cambridge: Cambridge University Press, 2000. Y por Spirtes, Glymour, y Scheines en: Causation, Prediction, and Search, 2nd ed., Cambridge: MIT Press, 2000.

<sup>68</sup>Hiddleston, Eric. "A causal theory of counterfactuals". Noûs, Vol. 39, No. 4 (Dec., 2005), p. 638-652.

la siguiente forma:

$$((Y_1 = y_1 \wedge \dots \wedge Y_n = y_n) \Rightarrow p(X = x) = z$$

Donde ‘ $\Rightarrow$ ’ es un condicional estricto,  $Y_i$  son todos los padres de  $X$ ,  $y_i$  son sus valores específicos, y  $z$  es una probabilidad “objetiva” de que  $X$  tenga el valor  $x$ .

Un *esquema modelo causal* es una dupla  $\langle G, E \rangle$ , donde  $G$  es una gráfica (una GAD) que consiste de un conjunto de variables  $V$  y flechas entre ellas; y  $E$  es un conjunto de ecuaciones relacionado a los valores (o probabilidades de valores) para cada variable  $X$  en  $G$  para los valores de sus padres  $pa(X)$  en  $G$ .

Un *modelo causal* es una tripleta  $\langle G, E, A \rangle$ , donde  $\langle G, E \rangle$  es un esquema modelo, y  $A$  es una asignación de valores para las variables en  $G$  lo cual es *posible* dado  $E$ . Esto es que, ninguna variable  $X$  tiene un valor  $A(X)$  tal que las ecuaciones digan que tiene probabilidad 0 dado los valores  $A$  asignado a los padres de  $X$ : para cualquier  $X$ ,  $p(A(X)|A(pa(X))) > 0$ .

Como ya se mencionó, las flechas dentro de los modelos representan relaciones de influencia causal. Las causas son tomadas como influencias positivas, mientras que los preventorios como influencias negativas. La relación entre las variables  $X \rightarrow Y$  es: el valor de  $X$  es causalmente relevante para las probabilidades de los valores de  $Y$ .

Sin embargo, para casos como el lanzamiento de la moneda, es necesaria la relación  *$x$  es una causa real de  $y$*  en nuestra teoría de contrafácticos. La intención de Hiddleston es evitar que su teoría sobre contrafácticos se encuentre atada a una teoría de causas reales, pero es necesario tener un concepto lo suficientemente preciso que nos dé predicciones definitivas acerca del valor de verdad de los contrafácticos. Por ello utilizará como sustituto el concepto de *influencia positiva directa*.

La propuesta consiste en usar la “influencia positiva directa” para caracterizar un “rompimiento causal” en un modelo, y después ofrecer un modelo “mínimamente alterado”. Así, un contrafáctico  $\phi > \psi$  es verdadero en un modelo  $M$  cuando  $\psi$  es verdadero en los  $\phi$ -modelos que son mínimamente alterados de  $M$ .

Supongamos  $M = \langle G, E, A \rangle$  es nuestro modelo “real”, y  $A_1, A_2, \dots$ , son las posibles



valoraciones alternativas para  $\langle G, E \rangle$ . Dejemos  $M_i = \langle G, E, A_i \rangle$ .

Supongamos que  $X$  es un padre de  $Y$  en  $M$ ,  $X = x$ ,  $Y = y$ , otros padres de  $Y$  son  $Z^\rightarrow$ , y esos padres tienen el valor  $z^\rightarrow$ .

**Influencia positiva directa** (primera aproximación):  $X = x$  tiene influencia positiva directa en  $Y = y$  en  $M$  si y sólo si  $p(Y=y \mid X=x \ \& \ Z^\rightarrow = z^\rightarrow) > p(Y=y \mid X \neq x \ \& \ Z^\rightarrow = z^\rightarrow)$

Dejemos que el padre positivo de  $Y$  en  $M$  sea  $ppa(Y)_M = \{X: X = x \text{ tiene influencia positiva directa en } Y = y \text{ en } M\}$ .

Esto es:  $X = x$  tiene influencia positiva directa en  $Y = y$  cuando el padre de  $X$  tiene el valor  $x$  *aumenta la probabilidad* de que el hijo de  $Y$  tenga valor  $y$  (dando los valores reales de los otros padres de  $Y$ ).  $ppa(Y)_M$  es el conjunto de padres positivos de  $Y$  en  $M$ .

**Rompimiento causal:** Un rompimiento causal en un modelo  $M_i$  relativo a  $M$  es una variable  $Y$  tal que  $A_i(Y) \neq A(Y)$ , y para cada  $X \in ppaM(Y)$ ,  $A_i(X) = A(X)$ . *Rompimiento*  $(M_i, M) = \{Y: Y \text{ es un rompimiento causal en } M_i \text{ a partir } M\}$ .

Esto es, un rompimiento causal (en  $M_i$  a partir de  $M$ ) ocurre en una variable  $Y$  que tiene un valor no real en  $M_i$  a pesar del hecho de que todos los padres de  $Y$  tienen sus valores reales en  $M_i$ . Si una variable  $Y$  no tiene padres, entonces las diferencias en los valores de  $Y$  cuenta como un rompimiento: mismos valores para todos los padres positivos, pero diferentes valores para  $Y$ . *Rompimiento*  $(M_i, M)$  es el conjunto de rompimientos de  $M_i$  a partir de  $M$ .

Además de estos elementos, necesitamos representar los casos donde los modelos preservan relaciones causales entre variables.

*Intacto*  $(M_i, M) = \{Y : A_i(Y) = A(Y) \text{ y para cada } X \in ppa(Y)_M, A_i(X) = A(X)\}$ . *Intacto* es el conjunto de variables  $Y$  para el cual  $M_i$  da valores reales tanto para  $Y$  como para los padres positivos de  $Y$ .

Para esclarecer lo anterior, consideraremos dos modelos del caso en el que Alice y Ben apuestan por fotones y que se ilustra en la Figura 2. Dentro del modelo 2(a), Ben apuesta por transmisión (Apuesta = t), el fotón se refleja, y Ben pierde. En el modelo 2(b), Ben apuesta

por reflexión (Apuesta = r) y gana. 2(b) contiene un rompimiento causal de 2(a) en Apuesta: mismos valores de padres positivos de Apuesta como en 2(a), es decir ninguno, diferente valor de Apuesta. Así tenemos que:  $Rompimiento(M_{2(b)}, M_{2(a)}) = \text{Apuesta}$ . Los padres de Emitir (ninguna) tienen el mismo valor en 2(a) y en 2(b), y también lo hace Emitir, de manera análoga para Reflejar.  $Intacto(M_{2(b)}, M_{2(a)}) = \text{Emitir, Reflejar}$ . Recordemos que el valor de Ganar difiere en los modelos, Ganar no cuenta como un rompimiento causal, los padres positivos de Ganar (Apostar) tienen diferentes en 2(a) y en 2(b), por lo que Ganar no es candidato para ser incluido en *Rompimiento* o *Intacto*.

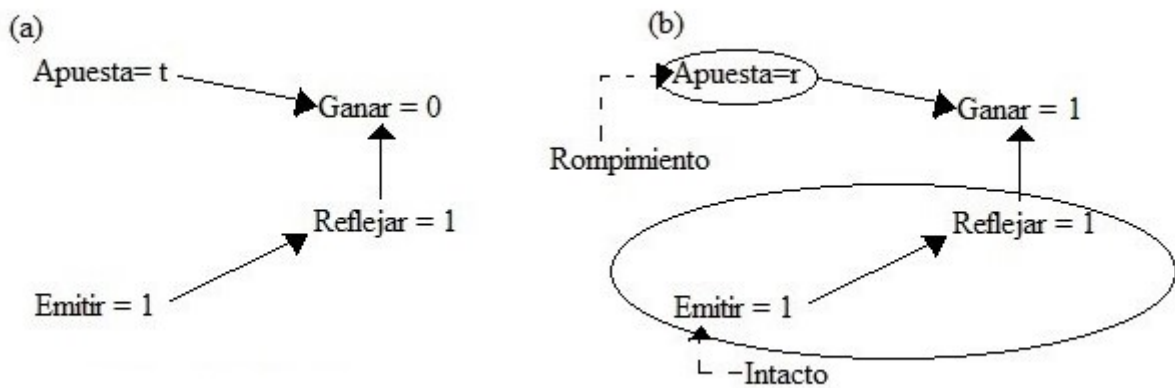


Figura 4: Apuesta por fotones

El siguiente paso es caracterizar un modelo “mínimamente alterado”. Digamos que un  $\phi$ -modelo es un modelo cuya proposición  $\phi$  es verdadera. *Rompimiento* e *Intacto* son conjuntos de variables, que serán más grandes, más pequeños o simplemente incomparables con otros conjuntos de variables. Diremos que  $Rompimiento(M_i, M)$  es mínimo entre  $\phi$ -modelos si y sólo si no hay un  $\phi$ -modelo  $M_k$  tal que  $Rompimiento(M_i, M) \supseteq Rompimiento(M_k, M)$ . Esto es  $Rompimiento(M_i, M)$  es mínimo (entre  $\phi$ -modelos) cuando ningún otro  $\phi$ -modelo incluye sólo un conjunto más pequeño de rompimientos en la relación causal de  $M$ .

De manera similar, diremos que  $Intacto(M_i, M)$  es máximo entre  $\phi$ -modelos si y sólo si no hay un  $\phi$ -modelo  $M_k$  tal que  $Intacto(M_i, M) \subset Intacto(M_k, M)$ . Esto es,  $Intacto(M_i, M)$  es máximo (entre  $\phi$ -modelos) cuando no haya otro  $\phi$ -modelo que comparta un conjunto “más grande” de relaciones causales intactas con  $M$ .

Para la siguiente definición vamos a suponer que  $\phi$  es una proposición atómica,  $X=x$ , o una negación o conjunción de ellas.

**Modelo  $\phi$ -mínimo:**

Modelo  $M_i$  y  $Rompimiento(M_i, M)$  son  $\phi$ -mínimos relativos a  $M$  si y sólo si:

- (I)  $M_i$  es un  $\phi$ -modelo,
- (II) para  $Z$  el conjunto de variables en  $G$  no hay descendientes de  $\phi$ ,  $Intacto(M_i, M) \cap Z$  es máximo entre  $\phi$ -modelos, y
- (III)  $Rompimiento(M_i, M)$  es mínimo entre  $\phi$ -modelos.

En esta teoría, los  $\phi$ -mundos más “similares” son sustituidos por los modelos  $\phi$ -mínimos que acabamos de definir. Recordemos que, si  $\phi$  es verdadero en  $M$ , entonces  $M$  es el único modelo  $\phi$ -mínimo relativo a  $M$  y  $\{\}$  es el único  $Rompimiento$  mínimo.

Nuestras cláusulas (II) y (III) especifican dos componentes de un “trasfondo fáctico” de  $\phi$ . La cláusula (II) nos dice que un trasfondo fáctico para  $\phi$  incluye un máximo conjunto de eventos actuales que no son causalmente influenciados por  $\phi$ . Los cambios a realizar deben ser lo más mínimos posibles y consistentes con que  $\phi$  sea verdadero, lo cual obliga a cualquier rompimiento causal a ser lo más tarde y menor que las leyes permitan. La cláusula (III) dice que un trasfondo causal también incluye un máximo conjunto de condicionales materiales del tipo: si los posibles padres de  $X$  tienen sus valores reales, entonces  $X$  tiene su valor real; nos dice que una relación real de causa-efecto se puede obtener si  $\phi$  fuera verdadera. Aproximadamente, las tres cláusulas juntas tienen el efecto de que  $\phi \dot{>} \psi$  es verdadero si y sólo si  $\psi$  se sigue legalmente de: (I)  $\phi$ , (II) eventos no causalmente influenciados por  $\phi$ , y (III) relaciones causales reales que  $\phi$  no previene de obtener.

**TCM (Teoría de Contrafácticos en un Modelo):**

$(\phi > \psi)$  es verdadero en un modelo  $M$  y en un contexto  $C$  si y sólo si  $\psi$  es verdadero en todo modelo  $\phi$ -mínimo  $M_i$  para el que  $Rompimiento(M_i, M)$  es relevante en  $C$ .

La cualidad de rompimiento “relevante” sirve para capturar algunos hechos acerca de la dependencia de contexto, aunque de momento los dejaremos de lado y regresaremos a ellos más adelante. Por ahora diremos que de acuerdo con TCM  $(\phi > \psi)$  es verdadero en  $M$  si y sólo si  $\psi$  es verdadero en todos los modelos que son  $\phi$ -mínimos en relación a  $M$ . Para ejemplificar TCM se utiliza el caso del fotón (Fig. 2). Supongamos que el modelo 2(a) representa el caso real. 2(b)

da el único modelo mínimo para el antecedente en el que Ben apuesta por reflexión (Apuesta =  $r$ ).  $Rompimiento(M_{2(b)}, M_{2(a)}) = \text{Apuesta}$ , y no hay un *Rompimiento* menor que sea consistente con  $\text{Apuesta} = r$ . De manera similar,  $\text{Intacto}(M_{2(b)}, M_{2(a)}) = \text{Emitir}$ , Reflejar es máximo en los modelos ( $\text{Apuesta} = r$ ). Para el antecedente  $\text{Apuesta} = r$ , “mantenemos fijos” los valores reales de Emitir y Reflejar: conservarían sus valores reales independientemente de las probabilidades previas. Añadimos un rompimiento mínimo a partir del modelo 2(a) con  $\text{Apuesta} = r$ , y vemos qué se sigue. El valor de Ganar “está en juego”, pero se sigue de las leyes del modelo que  $\text{Ganar} = 1$ . Por lo que es verdadero en el modelo 2(a) que Ben habría ganado si hubiera apostado a reflexión.

Ahora supongamos que Alice pudiera hacer trampa en el fotón reflector (Fig.3), ella tiene un interruptor de pie que cambia el ángulo de reflexión en la caja, haciendo que sea más probable que el fotón sea reflexivo (ahora la probabilidad cambiaría de 0.2 a 0.8). Ben apuesta por transmisión, Alice activa el interruptor de pie, y el fotón se refleja. Ahora parece fuertemente falso que “Si Ben hubiera apostado por reflexión, el habría ganado”.

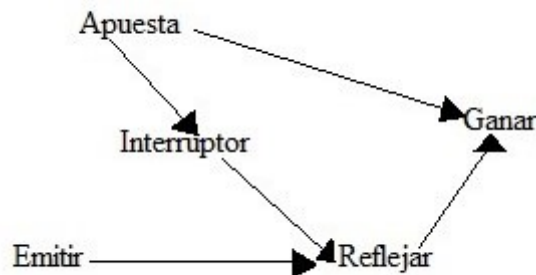


Figura 5: Trampa de fotones

Supongamos que las leyes de nuestro modelo  $M$  incluyen  $\text{Interruptor} = \text{Apuesta}$ . TCM recomienda que “ $\text{Apuesta} = r > \text{Ganar} = 1$ ” es falso en  $M$ . De nuevo, cualquier modelo en el que  $\text{Apuesta} = r$  incluye un rompimiento a partir de  $M$  en  $\text{Apuesta}$ . Pero el valor no real de  $\text{Apuesta}$  se propaga por el sistema:  $\text{Interruptor}$  no es un candidato para la inclusión de *Rompimiento* o *Intacto* porque su padre positivo  $\text{Apuesta}$  tiene un valor no real; pasa de manera similar con  $\text{Reflejar}$  y  $\text{Ganar}$ . El único máximo  $\text{Intacto}(M_i, M)$  es  $\text{Emitir}$ , y el único mínimo  $\text{Rompimiento}(M_i, M)$  es  $\text{Apuesta}$ . Los valores de las otras variables “están en juego”, y esta vez las leyes no son las únicas que determinan sus valores. Así que, “ $\text{Apuesta} = r > p(\text{Ganar}) = 0.2$ ” es verdadero, pero es falso tanto que Ben habría ganado, como que Ben habría perdido.

La Fig. 4 representa un caso de un cañón ceremonial que explota, creando un destello y un estallido. Supongamos que las variables son binarias y todas tienen valor 1 en  $M$ ; y supongamos que las leyes relevantes en  $M$  son:

a) (Fusible encendido = 1)  $\Rightarrow$  p(Explosión = 1) = 0.95,

b) Destello = Explosión

c) Estallido = Explosión

Esto es, el fusible es imperfecto, pero una vez que ha habido una explosión, definitivamente habrá un destello y un estallido. TCM cuenta como verdadero a “Si el destello no hubiera ocurrido, entonces el estallido no habría ocurrido”, “Destello = 0  $\wedge$  Estallido = 0”.  $Intacto(M_i, M) = Fusible\ encendido$  es máximo para Destellos = 0, porque no hay una asignación posible en la que Destello = 0 y Explosión = 1. Por ley (b) Destello = 0 necesita Explosión = 0, y por (c) necesita que Estallido = 0.

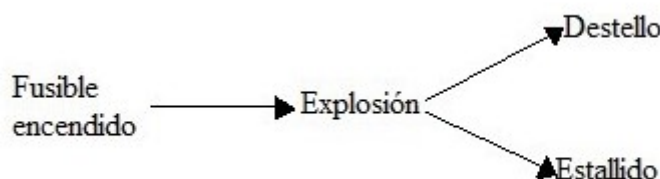


Figura 6: Un cañón.

En TCM “Destello = 0  $>$  Fusible encendido = 1” resulta verdadero: el fusible puede seguir estando encendido si el destello no ocurre. TCM permite y requiere “retroceder”, pero sólo cuando las causas de un evento sean legalmente necesarias. Esto es que TCM da una semántica unificada de “rastreadores”; además de explicar los “rastreadores”, también da una explicación de la problemática de los ejemplos 1-7, y también captura la idea de “mantener fijo el pasado” y dejar que el futuro actúe con respecto a las leyes reales.

Regresaremos a la cualidad que tiene TCM sobre *Rompimientos* contextualmente relevantes los cuales, a juicio de Hiddleston, sus contrafácticos tendrán importantes ventajas generales sobre los de Lewis.

El siguiente ejemplo es un ejemplo estándar de contexto-dependencia:

Xenon: Si saltaras por la ventana, estarías lesionada.

Yvonne: No, no lo estaría. Si tuviera que saltar, habría puesto una red de seguridad primero, y amortiguaría mi caída.

A primera instancia, no hay nada que nos haga decidir entre el contrafáctico enunciado por Xenon y el enunciado por Yvonne, sólo el contexto y nuestros propósitos momentáneos pueden decir cuál de ellos preferiríamos en una ocasión determinada.

La Fig. 5 ilustra un modelo natural del caso. Yvonne salta sólo si Red de Seguridad = 1 o Racional = 0. En el caso real, Red de Seguridad = 0 y Racional = 1, no salta (Salta = 0) y no está lesionada (Lesión = 0). La evaluación en TCM acerca de los *Rompimientos* contextualmente relevantes pretende capturar este caso: Xenon trata Racional como un Rompimiento mínimamente relevante, mientras Yvonne trata Red de Seguridad como relevante. Cada uno de estos son rompimientos (Salta =1)-mínimos, y son *incomparables*: cada uno renuncia a algo que el otro conserva. Las circunstancias de la pronunciación permiten a Xenon e Yvonne tomar decisiones diferentes.

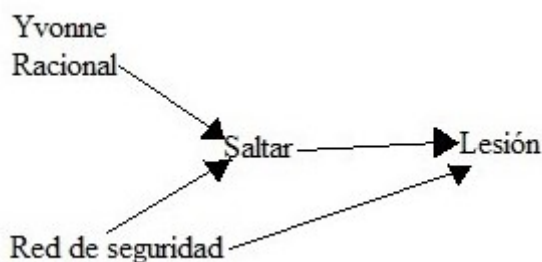


Figura 7: Dependencia contextual.

Existe una distinción de teorías, presentada por Lewis, a las que llama teorías “semánticas de premisa”, donde entraría la presentada por Hiddleston, en contraste con la suya a las que llama “semánticas de ordenamiento”. La idea general de las semánticas de premisa es que  $A > C$  es verdadero si y sólo si  $C$  se sigue legalmente de  $A$  y un conjunto de hechos reales  $F$ , un “trasfondo factual” de  $A$ . Estas teorías no son completamente diferentes, sin embargo, Hiddleston apunta al menos dos diferencias en las cuales la semántica de premisas es preferible.

La primera de esas diferencias a la que apunta es que la “similitud de mundos” es “global” y “aproximada”. En contraste, “minimalidad”<sup>69</sup> que es tanto local como “exacta”: sólo el

---

<sup>69</sup> *Minimality*.

acuerdo de modelos sobre un trasfondo fáctico relevante. De acuerdo con Lewis, la semejanza aproximada de mundos con respecto a eventos no necesitados del trasfondo factual hace una diferencia para las condiciones de verdad de los contrafácticos.

Sin embargo, esto no es así. Supongamos que Nixon está sentado mirando a Jane, mientras su dedo se encuentra en el botón, dispuesto a oprimirlo si ella lanzara la moneda en su mano y cayera cara. Ella gasta la moneda sin haberla lanzado, y Nixon de mala gana pone a un lado el botón. Si hay *cualquier* contexto en el que la similitud de mundos posibles, global y aproximada, cuenta para *algo*, sería en *algún* contexto en el que es verdadero “Si Jane hubiera lanzado la moneda, habría caído cara”. Pero no hay tal contexto. De acuerdo con TCM, no hay manera en que el contexto pueda forzar a un resultado no real de un proceso causal no real.

La segunda diferencia la consideración de Hiddleston de modelos  $\phi$ -mínimos permiten *incomparabilidad* generalizada. En contraste, la relación de “similitud” en Lewis es un *ordenamiento* de mundos posibles, y está “conectado”: para cuales quiera dos mundos son más, menos o igualmente similares al actual.

La Fig. 6 es un ejemplo de John Pollock<sup>70</sup> en el cual Hiddleston intenta ilustrar que su propuesta.  $L_i$  son luces, y  $S_i$  son interruptores. En el caso real, los interruptores están abiertos ( $\neg S_i$ ) y las luces están apagadas ( $\neg L_i$ ). Los interruptores funcionan de la siguiente manera:

- $L_1 = S_1 \vee S_2$
- $L_2 = S_2 \vee S_3$

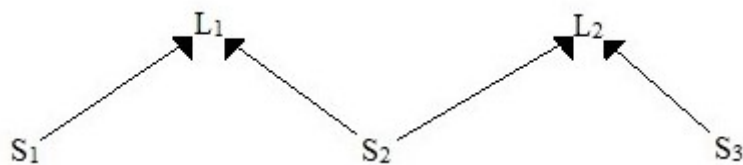


Figura 8: Caso de Pollock.

Esto es, una luz está encendida si y sólo si al menos uno de los interruptores conectados a él está cerrado.

<sup>70</sup>Pollock, John. “A refined theory of counterfactuals”. *The journal of philosophical logic* 10, 1981.

La teoría de Lewis valida el siguiente principio:

Fortalecimiento Antecedente Limitado (FAL):

(I)  $P > Q$

(II)  $\neg(P > \neg R)$  implica

(III)  $(P \wedge R) > Q$ .

Supongamos (I) y (II) son verdaderos, y consideremos los  $P$ -mundos más similares. De (i), son todos  $Q$ -mundos. De (II), algunos de ellos son  $R$ -mundos. (III) nos pide buscar los  $(P \wedge R)$ -mundos más similares de (I) en los que  $R$  también se sostiene. Pero  $Q$  se sostiene a través de los  $P$ -mundos más similares. Por lo que (III) es verdadero en Lewis.

FAL se basa en la conectividad, pero esto es intuitivamente inválido. En el caso de los interruptores, los siguientes contrafácticos son intuitivamente correctos:

- (i')  $L_2 > \neg S_1$
- (ii')  $\neg(L_2 > \neg L_1)$
- (iii')  $\neg[(L_2 \wedge L_1) > \neg S_1]$ .

(iii') es la negación de Una instancia de (iii), para  $Q = \neg S_1$ .

TCM toma a (i')-(iii') como verdaderos, y no valida LAS. La razón de esto es porque un rompimiento en  $S_2$  es incomparable con un rompimiento en  $S_1$  y  $S_3$ : Si  $L_2 \wedge L_1$  hubieran sido, entonces eso podría haber sido por  $S_2$ , pero igual podría haber sido por  $S_1$  y  $S_3$ . Tanto  $S_2$  como  $S_1$ ,  $S_3$  son rompimientos  $(L_1 \wedge L_2)$ -mínimos: cada uno nos da algún hecho real que el otro guarda. Por lo que TCM dice que no es el caso que si  $(L_1 \wedge L_2)$ , entonces podría haber sido por  $\neg S_1$ ; pero esto es verdadero porque podría haber sido que  $\neg S_1$ .

La explicación general de Hiddleston sobre contrafácticos es que son verdaderos *simpli-citer* cuando son verdaderos en buenos modelos.

Vamos a decir que un “caso” es un conjunto de eventos. Trataremos a un evento como un objeto (/n-tuplas) que tiene propiedad (/relación) en un tiempo o durante un tiempo. Hiddleston asume que las variables en modelos representan propiedades en la primera instancia



(variables de propiedad, realmente), y las ecuaciones representan legalmente relaciones entre propiedades.

### **Buen modelo:**

$M$  es un buen modelo para un caso  $C$  si y sólo si:

- (I) Las propiedades representadas en  $M$  son instanciadas por objetos en  $C$ ,
- (II) Las leyes de  $M$  son exactas de  $C$ , o suficientemente exactas, y
- (III)  $M$  es suficientemente completo para representar precisamente las relaciones entre los eventos de  $C$  que aparecen en  $M$ .

Esta sugerencia está supuesta en función de lo siguiente:

### **TC** (Teoría de contrafácticos):

$A > B$  es verdadero del caso si y sólo si hay un buen modelo de  $C$ , y  $A > B$  es verdadero en  $M$ .

Verdad-en-un-modelo es dada por TCM.

La cláusula (I) de “Buen Modelo” requiere simplemente (por ejemplo) que si un modelo tiene una variable por si una moneda es lanzada y toma un valor 1, entonces el caso incluye una moneda lanzada. Clausulas (II) y (III) requieren explicación adicional.

La cláusula (III) es el intento de dar la teoría de los contrafácticos mientras se intenta evitar los detalles de una teoría completa de la causalidad. No se apela al concepto de una “causa real”, pero en su lugar se usa como sustituto el concepto de “influencia directa positiva”. Existen casos bien conocidos en que las causas y la influencia positiva son diferentes, lo cual plantea problemas para la teoría de contrafácticos de Hiddleston, y dichos problemas provienen de la divergencia del concepto sustituto de la de una causa real.

Fig. 7 ilustra un caso de “derecho de prioridad”. Dos asesinos conspiran para matar un rey: el Asesino Maestro y el Asesino Torpe. El Asesino Maestro escondido en el clóset, con una daga en la capa. El Asesino Torpe mezcla veneno en la sopa del rey, y la lleva desde la cocina. Pero el Asesino Torpe en verdad es un asesino torpe, y él tropieza con la puerta del closet cuando pasa por ella, encerrando al Asesino Maestro dentro. Después se encuentra frente a una silla.

Tropieza y la sopa se derrama en el tazón, pero la recupera, y pone la sopa delante del rey, que se la come y muere poco después.

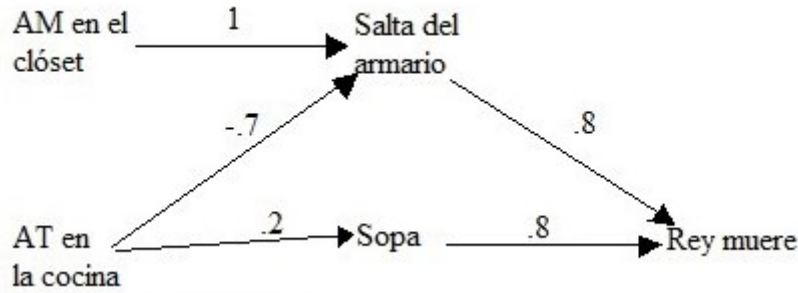


Figura 9: Derecho de prioridad.

TCM aplicado a este modelo obtiene las respuestas correctas para el caso. Por ejemplo,  $AT = 0 > RM = 1$  es falso. Podemos también representar el caso en la Figura 8. Dando reglas similares a estas asumimos el caso original,  $AM$  cuenta como una “influencia directa positiva” sobre  $RM$  en un modelo para Fig. 8, mientras que  $AT$  no. Así, TCM trata el caso similar a como lo hace en la Fig. 1, obteniendo la respuesta equivocada donde  $AT = 0 > RM = 1$  es verdadero.

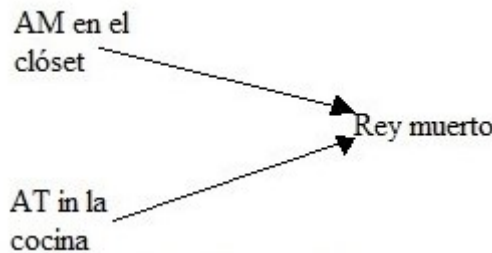


Figura 10: Derecho de prioridad.

Este ejemplo ilustra tanto que las influencias positivas ( $AM$ ) pueden fallar en ser causas reales, como que las influencias negativas ( $AT$ ) pueden ser. Hiddleston sugiere que todos los casos de influencias negativas que son causas, son casos en los que la causa interactúa con otras influencias en el efecto. Estas interacciones ocurren en procesos de intervención causal. Para todos los casos de influencias positivas que no son causas reales son casos en que algún proceso de intervención causal no se completa.

La cláusula (III) dice que  $M$  es “suficientemente completo”, lo que quiere decir que el modelo representa casos de influencia mixta como tal, e incluye variables suficientes para capturar si el proceso de intervención causal es completo. Esto nos señala que debemos preferir la Fig. 7 sobre 8 para el caso de los asesinos, lo que resolvería el problema anteriormente mencionado.

Uno de los temas más complejos de la teoría es proveer una explicación de los “padres positivos” que es sensible a los detalles de las influencias contrastivas. Supongamos que tenemos un modelo simple  $P \rightarrow D$ .  $P$  es una dosis de veneno, y toma el valor 0 (no veneno), 1 (dosis baja), y 2 (dosis alta);  $D$  es muerte. Supongamos por motivos de ilustración que el veneno es (al menos casi-) indeterminado.  $p(D|P = 0) = 0.01$ ,  $p(D|P = 1) = 0.5$ ,  $p(D|P = 2) = 0.99$ . La rata Fluffy tuvo una dosis baja, y sobrevivió. Tomaremos que “Si Fluffy no hubiera recibido ningún veneno, entonces habría sobrevivido” es correcto incluso aunque pensando que había un 0.01 de probabilidad de que muriera. Pero los dos siguientes resultan falsos: “Si Fluffy hubiera recibido una dosis alta, entonces habría sobrevivido”. “Si Fluffy hubiera recibido una dosis alta, entonces no habría sobrevivido”.

Necesitamos dar cuenta de influencias positivas que capturen estos hechos contrastantes. Supongamos que  $X$  es un padre de  $Y$  en  $M$ . Otros padres de  $Y$  son  $Z^{\rightarrow}$ , y  $Z^{\rightarrow} = z^{\rightarrow}$  en  $M$ .

### **Influencia positiva:**

$X = x$  en comparación con  $X = x'$  tiene influencia directa positiva en  $Y = y$  en  $M$  si y sólo si  $p(Y = y|X = x \wedge Z^{\rightarrow} = z^{\rightarrow}) > P(Y = y|X = x' \wedge Z^{\rightarrow} = z^{\rightarrow})$ .

Sea  $ppa(Y)_M, M_i = \{X : A(X)$  en comparación con  $A_i(X)$  tiene una influencia positiva directa en  $Y = y$  en  $M\}$ .

Esta explicación relativiza ser un padre positivo a un par de modelos.

Según Hiddleston la teoría TC es preferible sobre otras por su habilidad de explicar casos y brindar una explicación unificada de “rastreadores”. La moral filosófica de TC consiste es que reducir el análisis de contrafácticos de la causalidad es equivocado. Hay algo intuitivamente natural en esas teorías y TC explica qué es: entre los generadores estándar de verdad de un contrafáctico del tipo “Si  $C$  no hubiera ocurrido, entonces  $E$  no habría ocurrido” está el hecho de que  $C$  causa a  $E$ .

Una de las objeciones que podrían darse a TC es que la teoría de Lewis permanece correcta en los contextos estándar, mientras que los ejemplos 1-7 involucran algún contexto no estándar. Para Hiddleston, ambas teorías coinciden en un rango amplio de casos, la distinción se encuentra en el planteamiento de las condiciones de verdad: en Lewis la idea es mantener

fijo el pasado del antecedente  $A$ , añadir divergencias mínimas que nos permitan que  $A$  sea verdadero, dejar que el futuro siga como las leyes lo permiten,  $A > C$  es verdadero si y sólo si  $C$  también lo es; la idea de Hiddleston es similar con excepción de que lo que se mantiene fijo no es el pasado del antecedente, sino los hechos reales que son *causalmente independientes* del antecedente. Así, aunque ambas teorías coinciden en muchas de sus predicciones, en los casos en los que desacuerdan TC es preferible.

## Conclusiones

Como se mencionó en un principio, Hiddleston busca desarrollar una teoría distinta a la propuesta por Stalnaker-Lewis, a partir de modelos causales, para así abarcar más casos respecto a lo que las semánticas de mundos posibles alcanzan a cubrir. Este apartado tiene como objetivo señalar las diferencias y similitudes que puedan existir entre las teorías, y señalar de manera general las virtudes y ventajas que podrían llegar a tener una sobre la otra.

Lo primero que Hiddleston hace para atacar la teoría Stalnaker-Lewis es apelar a las condiciones de verdad que se proponen para los contrafácticos y a la resolución estándar de vaguedad a la que refiere para cualquier similitud. Por una parte, pareciera que Hiddleston no reproduce de manera justa dichas condiciones de verdad: en el apartado 2.2 del presente trabajo se expuso la propuesta de Lewis en la cual, lo primero que se hace es distinguir entre dos condicionales contrafácticos, el ‘might’ y el ‘would’. Las condiciones que Hiddleston señala y que presenté al principio del apartado 3.1 parecen apuntar, de manera vaga, a las condiciones de verdad del contrafáctico ‘might’.

La objeción de Hiddleston se encuentra basada en el rechazo a utilizar la similitud de mundos y la vaguedad que ésta representa, para atacar esta idea nos presenta casos que desafían los alcances de la teoría Stalnaker-Lewis. Sin embargo, me parece que algunos de estos ejemplos podrían llegar a resultar un tanto problemáticos ya que parecen entrañar la idea de que se tiene cierto acceso epistémico privilegiado. Para aclarar esto último hablaré de ejemplo 4 presentado en 3.1; en dicho ejemplo Alice tiene la capacidad de influir el resultado del lanzamiento de la moneda, por lo que asumir que Alice hace que la moneda caiga cara o cruz significaría asumir que de cierta manera podemos tener conocimiento de sus acciones incluso antes de que las realice.

Con los señalamientos anteriores lo que quiero decir no es que la teoría de Hiddleston sea equivocada, mi única intención es señalar algunas de las objeciones que se podrían tener, y a pesar de éstas, la teoría también guarda algunas virtudes. Una de ellas es que Hiddleston retoma la idea intuitiva de que los condicionales contrafácticos guardan algún tipo de conexión causal entre el antecedente y el consecuente, para ello hace uso de algunos términos como el de “ruptura causal”, “influencia directa positiva” y “modelo  $\phi$ -mínimo”.

Para Hiddleston, sus “modelos  $\phi$ -mínimos” tienen ventaja sobre la “similitud de mundos” dado que el primero atiende a un trasfondo fáctico relevante y es más exacto, mientras que lo segundo nos da algo más general. Aunque él lo ve como una virtud, esto también podría resultar problemático, dado que si la similitud de mundos es más general abarcaría más casos de los que podría abarcar los modelos  $\phi$ -mínimos. De igual modo los modelos  $\phi$ -mínimos pueden parecer menos artificiales por que apelan al trasfondo fáctico, mientras que la similitud no. Otra de las ventajas que Hiddleston señala sobre los modelos  $\phi$ -mínimos es que permiten *incomparabilidad*, mientras que en la similitud debe haber un ordenamiento de los mundos, por lo que la teoría de Hiddleston podría dar solución a decidir sobre contrafácticos de la forma  $A > B$  y  $A > \neg B$ , donde la teoría Stalnaker-Lewis parece no poder decidir.

Hiddleston señala repetidas veces que su teoría es preferible porque dan respuesta a contrafácticos que involucran algún tipo de contexto no estándar, lo cual podría parecer verdad, sin embargo, me parece importante delimitar qué es lo que queremos cuando buscamos una teoría sobre contrafácticos, si lo que queremos es a) tener una teoría general que nos permita abarcar todas las enunciaciones contrafácticas; b) abarcar solamente ciertos casos específicos; o c) hacer una taxonomía de los diferentes tipos de contrafácticos y presentar una evaluación distinta para cada uno de ellos. Si nuestras intenciones son a), se trata de algo que, al menos parece, no ser posible, se han propuesto muchas teorías que intenten dar cuenta de esto y que han fracasado; sin embargo, considero que la teoría Stalnaker-Lewis es lo que podría acercarnos de forma satisfactoria. Si nuestras intenciones son b), va a depender de cuáles son los casos que queremos abarcar, y en alguna medida la teoría de Hiddleston podría proporcionar soluciones a nuestros fines. Pensar en c) nos llevaría a hacer un trabajo un tanto distinto al presentado aquí, ya que primeramente se necesita hacer la clasificación de los tipos de contrafácticos que pensamos que hay para después proponer para cada uno de ellos una evaluación; no obstante, intentar clasificar los contrafácticos podría resultar problemático, ya que cabe la posibilidad de que distinción entre uno y otro tipo de contrafáctico no sea clara.

Aunque parezca que no somos capaces de tener una teoría general para la evaluación de condicionales contrafácticos, me parece que la teoría Stalnaker-Lewis es suficientemente satisfactoria para abarcar un buen número de casos. A nivel personal me interesa el desarrollo de un trabajo que involucre la distinción entre tipos de contrafácticos —a pesar de lo problemático que dijimos que podría resultar— y así contrastar las distintas estructuras que un contrafáctico puede tener y si es el caso o no que funcionaría mejor que una teoría general o no.

## Referencias

- [1] Adams, E. "Subjunctive and indicative conditionals." *Formulations of language* 6, pp. 89-94, 1970.
- [2] Baker, Stephen. "Counterfactuals, probabilistic counterfactuals, and causation". *Mind* 108, 1999.
- [3] Bennet, Jonathan. "Meaning and implication". *Mind*. Vol. 63, No. 252, p. 451-463, 1954.
- [4] Djordjevic, Vladan. "similarity and cotenability". *Synthese* 190, p 681-691, 2013.
- [5] Edgington, Dorothy. "On Conditionals". *Mind* 104, 235. 1995.
- [6] Gómez Torrente, Mario. "La teoría semántica de las lógicas de condicionales de Stalnaker y Lewis." *En Agora* 12/2 p. 57-78, 1993. P. 67.
- [7] Goodman, Nelson. "The problem of counterfactual conditionals". *The Journal of Philosophy*, Vol. 44, No. 5, p. 113-128, 1947.
- [8] Goodman, Nelson. *The structure of Appearance*, Cambridge, Harvard University Press, 1951.
- [9] Heylen Jan & Horsen Leon. "Strict conditionals: A negative result". *The Philosophical Quarterly*, Vol. 56, No. 225, 2006.
- [10] Hiddleston, Eric. "A causal theory of counterfactuals". *Noûs*, Vol. 39, No. 4 (Dec., 2005), pp. 632-657.
- [11] Jackson, Frank. *Conditionals*. Oxford: Basil Blackwell. 1987.
- [12] Kripke, S. "Semantical analysis of modal logics", I, *Zeitschrift für mathematische Logik and Grundlagen der Mathematik*, vol. 9, p. 67-96, 1963.
- [13] Kvart, Igal, "Causal Independence". *Philosophy of Science* 61, 96, 1994.
- [14] Kvart, Igal. "Transitivity and Preemption of Causal Relevance". *Philosophical Studies* 64, p. 125-60, 1991.
- [15] Kvart, Igal. *A Theory of Counterfactuals*. Indianapolis: Hackett. 1986.
- [16] Lewis, C. I. & Langford C. H., *Symbolic Logic*, New York, 1932.

- [17] Lewis, C. I., Survey of symbolic logic, University of California press, Berkeley, 1918.
- [18] Lewis, C.I. "Implication and the algebra of logic". En Mind, Vol. 21, No. 84 (Oct., 1912), pp. 522-531.
- [19] Lewis, David. Counterfactuals. Cambridge: Harvard University Press, 1973.
- [20] Morado, Raymundo. Instituto de Investigaciones Filosóficas, UNAM. 20 de Marzo de 2018 "Implicación Material", <http://www.filosoficas.unam.mx/morado/Cursos/11FilLog3Sem/110901.pdf>
- [21] Palau, G. Condicionales contrafácticos: condiciones de verdad y semántica de mundos posibles. Acerca de las teorías de R. Stalnaker y D. Lewis.<sup>en</sup> Critica, vol.XII No. 34, México. 1980.
- [22] Pearl, Judea, "Probabilities of Causation: Three Counterfactual Interpretations and Their Identification". Synthese 121, 93-149, 1999.
- [23] Pearl, Judea. Causality. Cambridge: Cambridge University Press, 2000.
- [24] Pollock, John. "A refined theory of counterfactuals". The journal of philosophical logic 10, 1981.
- [25] Ramsey, "General propositions and causality", en Foundations, Londres, 1978. p. 143.
- [26] Russell B. & Whitehead, A.N. Principia mathematica, Cambridge, 1910-1913, 3 vols.
- [27] Sexto Empírico, Esbozos pirrónicos. Gredos, Madrid, 1993.
- [28] Spirtes, Peter, Clark Glymour, and Richard Scheines. Causation, Prediction, and Search, 2nd ed., Cambridge: MIT Press. 2000.
- [29] Stalnaker, R. & Thomason, R. H. "A semantic analysis of conditional logic", en Theoria 36 (1), p. 23-42, 1970.
- [30] Stalnaker, Robert. "A Theory of Conditionals". En N. Rescher ed., Studies in Logical Theory. Oxford: Basil Blackwell, 1968. pp. 98-112.