Update Semantics

Frank Veltman

Mini course at the Department of Philosophy, Peking University

Introduction

Semantics, the standard view

Starting Point: An argument is **valid** iff in every possible situation in which the premises of the argument are true, the conclusion is true as well.

Meaning: You know the **meaning** of a sentence iff you know its truth conditions.

Framework Possible world semantics.

Dynamic semantics

Meaning: You know the **meaning** of a sentence if you know the change it brings about in the intentional state of anyone who wants to incorporate the information conveyed by it.

- The meaning $[\varphi]$ of a sentence φ is an operation on intentional states.
- Let S be an intentional state and φ a sentence with meaning $[\varphi]$. We write

$$S[\varphi]$$

for the state that results when S is updated with φ.

Key notions

Support Sometimes the information conveyed by φ will already be subsumed by S. In this case, we say that φ is accepted in S, or that S supports φ, and we write this as $S \models \varphi$. In simple cases this relation can be defined as follows:

- $S \models \varphi$ iff $S[\varphi] = S$

Logical validity An argument is valid if updating any state with the premises, yields a state that supports the conclusion.

- $\varphi_1, \ldots, \varphi_n \models \psi$ iff for every state S, $S[\varphi_1] \ldots [\varphi_n] \models \psi$.

Problem: anaphoric reference

- I lost ten marbles, and found all but one of them. Probably it is under the sofa.

- I lost ten marbles and found nine of them. Probably it is under the sofa. (??)

- I lost ten marbles and found nine of them. Probably the missing one is under the sofa.

- I lost ten marbles, and found all but two of them. One probably is under the sofa, the other might be anywhere.

Problem: Epistemic Modalities

Compare:

- *Somebody is knocking on the door. . . It might be the milkman . . . It's Mary.*

- *Somebody is knocking on the door. . . It's Mary. . . It might be the milkman.*(??)

Or think of somebody saying:

- *John didn't drink too much wine. He would have got sick.*

- *John drank too much wine. He would not have got sick.* (??).

Puzzle

Consider these two dots:

● ●

One is called a, the other one is called b. I am not telling you which is which. Imagine me pointing at the red dot, saying "this". When I do so I will write " this$_{\bullet}$". Similarly for the blue bullet.

We will assume that the domain of quantification consists of just these two dots. So, the information you have right now is this:

- $\text{this}_{\color{red}\bullet} \neq \text{this}_{\color{blue}\bullet}$, $\forall x(x = \text{this}_{\color{red}\bullet} \vee x = \text{this}_{\color{blue}\bullet})$, $a \neq b$.

Now, one could reason as follows:

- $\mathit{might}(\text{this}_{\color{red}\bullet} = a) \wedge \mathit{might}(\text{this}_{\color{blue}\bullet} = a)$.
  So, $\forall x \mathit{might}(x = a)$,
  Therefore, $\mathit{might}(b = a)$.

Or:

- $\mathit{might}(a = \text{this}_{\color{red}\bullet}) \wedge \mathit{might}(b = \text{this}_{\color{red}\bullet})$.
  So, $\forall x \mathit{might}(x = \text{this}_{\color{red}\bullet})$.
  Therefore, $\mathit{might}(\text{this}_{\color{blue}\bullet} = \text{this}_{\color{red}\bullet})$.

Where are the mistakes?

Presupposition

In general, the update operation will be partial rather than total. Take the case of *definite descriptions*. Example:

The present king of France is bald.

One cannot update ones state with this sentence if one does not believe that France is a monarchy. The sentence presupposes that there exists a unique king of France.

Definition

φ *presupposes* ψ iff for every S, $S[\varphi]$ is defined only if $S \models \psi$.

See for example

Beaver, David, 2001, *Presupposition and assertion in dynamic semantics*. Vol. 29. Stanford: CSLI publications.

Schlenker, Philippe, 2007, ‘Anti-dynamics: Presupposition projection without dynamic semantics.’ *Journal of Logic, Language and Information*, 16, 325-356.

Rothschild, Daniel, 2011, ‘Explaining presupposition projection with dynamic semantics’, *Semantics and Pragmatics*, 4, 3-1

Problem: Paradoxes

Consider the **Principle of Idempotence**

For any state S and sentence φ, $S[\varphi] \models \varphi$

At first sight this principle goes without saying. But there are sentences for which it doesn't hold. Here paradoxical sentences like

This sentence is false

are a point in case:

There is no state S such that

$$S[\textit{This sentence is false}] = S$$

There is no such thing as a *successful* update with the Liar, and that is where its paradoxicality resides.

For more details, see:
Groeneveld, Willem, 1994), Dynamic Semantics and Circular Propositions, *Journal of Philosophical Logic*, 23, 267-306.

Static versus Dynamic Semantics

Seth Yalcin and Daniel Rotschild proved that if the update operation [ ] is idempotent and commutative, the semantics is essentially static.*

*Rothschild, D., & Yalcin, S. (2017). On the dynamics of conversation. *Noûs*, **51**, 24-48.

Might

Cognitive states

Definition 1
Fix a finite set $\mathcal{A}$ of atomic sentences.

(i) A *possible world* is a function with domain $\mathcal{A}$ and range $\{0, 1\}$;

(ii) $\mathcal{W}$ is the set of all possible woirlds;

(iii) S is an *information state* iff $S \subseteq \mathcal{W}$;

(iv) The *minimal state* is the information state given by $\mathcal{W}$; The *absurd state*, is the information state given by the empty set.

Updates

Definition 2

For every sentence φ and state S, $S[\varphi]$ is determined as follows:

$$\begin{aligned}
\textit{atoms}\text{:} \quad & S[p] = S \cap \{w \in \mathcal{W} \mid w(p) = 1\} \\
\neg : \quad & S[\neg\varphi] = S \sim S[\varphi] \\
\wedge : \quad & S[\varphi \wedge \psi] = S[\varphi] \cap S[\psi] \\
\vee : \quad & S[\varphi \vee \psi] = S[\varphi] \cup S[\psi]
\end{aligned}$$

Might

Basic idea

One has to agree to *might* φ if φ is consistent with ones knowledge — or rather with what one takes to be ones knowledge. Otherwise *might* φ is to be rejected.

Definition 3

$$\begin{array}{ll} \textit{might}: & S[\textit{might}\,\varphi] = S \text{ if } S[\varphi] \neq \emptyset \\ & S[\textit{might}\,\varphi] = \emptyset \text{ if } S[\varphi] = \emptyset \end{array}$$

Sentences of the form *might* φ provide an invitation to perform a test on S rather than to incorporate some new information in it.

Notice:

$$\mathcal{W}[\mathit{might}\, p][\neg p] \neq \emptyset;$$

$$\mathcal{W}[\neg p][\mathit{might}\, p] = \emptyset.$$

Some facts

Proposition 1 *Let* φ *be a sentence in which might does not occur. Set* $[\![\varphi]\!] = \mathcal{W}[\varphi]$ *Then for any information state* S*,* $S[\varphi] = S \cap [\![\varphi]\!]$*.*

If φ is a sentence of ordinary propositional logic, we can speak of 'the φ-worlds', the worlds in which the proposition expressed by φ holds. It would be nonsense to speak of the '*might* φ-worlds'. Sentences of the form '*might* φ' do not express a proposition.

Proposition 2 *Let* $\varphi_1, \ldots, \varphi_n, \psi$ *be sentences in which* might *does not occur.* $\varphi_1, \ldots, \varphi_n \models \psi$ *iff* $\varphi_1, \ldots, \varphi_n / \psi$ *valid in classical logic*

Proposition 3 *For the full language, the logic is non-monotonic.*[*]

Example

$$\textit{might}\,\neg p \models \textit{might}\,\neg p, \text{ but } \textit{might}\,\neg p, p \not\models \textit{might}\,\neg p$$

[*]If you are interested in a complete proof system for this logic, see: Does, J. van der, W. Groeneveld, and F.Veltman, 1997, 'An update on Might', *Journal of Language, Logic, and Information*, 6, 361-380.

Shortcoming?

'Might' does more than the test semantics suggests. With 'might' one can draw somebodies attention to certain possibilities, make somebody become aware of some possibilities, express that one has become aware of a certain possibiliy, . . .

Relevant literature

- Ciardelli, Ivano, Jeroen Groenendijk, and Floris Roelofsen. (2009). Attention!'Might'in inquisitive semantics. *Semantics and linguistic theory,* 91-108.
- Yalcin, Seth. (2011). Nonfactualism about epistemic modals. In: A. Egan and B. Weatherson (Eds.), *Epistemic modality*, Oxford University Press, 295-332.
- Willer, Malte. (2013). Dynamics of epistemic modality. *Philosophical Review*, 122(1), 45-92.

Indicative conditionals

Reminder: material implication and its paradoxes

p	q	$p \to q$
1	1	1
1	0	0
0	1	1
0	0	1

(i) $\neg p \models p \to q$

(ii) $q \models p \to q$

or, equivalently:

(iii) $\neg(p \to q) \models p$

(iv) $\neg(p \to q) \models \neg q$

Counterexamples

This way, we would have to accept that

- (i) *If* $2+2=5$, *Shanghai is the capital of the China.*
- (ii) *If Oswald killed Kennedy, Amsterdam is the capital of the Netherlands.*

What many logicians/linguists/philosophers find unsettling about these sentences is that in each of them the antecedent seems irrelevant to the consequent.

(iii) and (iv) are also counterintuitive, but here 'relevance' is not the issue.

- (iii)
  premise: *It's not the case that if the peace treaty is signed, war will be prevented.*

  ---
  conclusion: *The peace treaty will be signed,*

- (iv)
  premise: *It's not the case that if the peace treaty is signed, war will be prevented.*

  ---
  conclusion: *War will not be prevented,*

Similarly?

premise 1:	*If Jones wins the election, Smith will retire to private life.*
premise 2:	*If Smith dies before the election, Jones will win it.*
conclusion:	*If Smith dies before the election, he will retire to private life.*

Says A: "This is a clear-cut counterexample to a putative logical principle."

Says B: "This is just a pragmatically incorrect instance of a semantically faultless rule."

If. . . , then. . .

$\varphi \Rightarrow \psi$ gets an epistemic interpretation*, just like *might* φ.

Definition 4

$$\Rightarrow: \quad \begin{array}{l} S[\varphi \Rightarrow \psi] = S \textit{ if } S[\varphi] \models \psi \\ S[\varphi \Rightarrow \psi] = \emptyset \textit{ if } S[\varphi] \not\models \psi \end{array}$$

*Proposed in:
Gillies, A., 'Epistemic Conditionals and Conditional Epistemics'. In: *Noûs*, vol 38, 2004, pp. 585-615.

Logical properties

Modus Ponens is valid:

$$\varphi \Rightarrow \psi, \varphi \models \psi$$

The Deduction Principle holds, too:

$$\varphi_1, \ldots, \varphi_n \models \psi \Rightarrow \chi \text{ iff } \varphi_1, \ldots, \varphi_n, \psi \models \chi$$

Exercise

Prove that in the *standard framework* only material implication has these properties.

(In the "standard" framework (i) '$\models$' stands for the standard notion of validity in terms of truth. What is also needed as standard is (ii) the principle of bivalence, and as a natural consequence of this the assumption that (iii) the negation of a sentence is true iff the sentence itself is false).

Nevertheless

$$\neg(p \Rightarrow q) \not\models p \wedge \neg q$$

Instead we have:

$$\neg(p \Rightarrow q) \models \textit{might}\,(p \wedge \neg q)$$

However, we did not get rid of:

$$\neg p \models p \Rightarrow q$$

$$q \models p \Rightarrow q$$

Modus Tollens

The case of the marbles There are three marbles: one red, one blue, and one yellow. They are known to be distributed among two matchboxes, called 1 and 2. The only other thing which you are told is that there is at least one marble in each box.

Do you agree that

(1) *If the red marble is in box 2, then if the blue one is in box 2 as well, the yellow one is in box 1.*

(2) *It is not the case that if the blue marble is in box 2, the yellow one is in box 1.*

Given your information,
there are six possibilities left:

w_1	$b_1,\ r_2,\ y_2$
w_2	$y_1,\ b_2,\ r_2$
w_3	$r_1,\ b_2,\ y_2$
w_4	$y_1,\ r_1,\ b_2$
w_5	$r_1,\ b_1,\ y_2$
w_6	$b_1,\ y_1,\ r_2$

(i) S[*blue in 2*] $= \{w_2, w_4\}$

(ii) S[*blue in 2*][*yellow in 1*] $= \{w_4\}$

(iii) (i) and (ii) imply S[*if blue in 2, then yellow in 1*] $= \emptyset$, and therefore $S[\neg$(*if blue in 2, then yellow in 1)*] $= S$.

(iv) Hence, $S \models \neg$(*if blue in 2, then yellow in 1*)

w_1	$b_1,\ r_2,\ y_2$
w_2	$y_1,\ b_2,\ r_2$
w_3	$r_1,\ b_2,\ y_2$
w_4	$y_1,\ r_1,\ b_2$
w_5	$r_1,\ b_1,\ y_2$
w_6	$b_1,\ y_1,\ r_2$

(i) S[*red in 2*] $= \{w_1, w_2\}$

(ii) S[*red in 2*][*blue in 2*] $= \{w_2\}$

(iii) S[*red in 2*][*blue in 2*][*yellow in 1*] $= \{w_2\}$

(iv) So, S[*red in 2*][*if blue in 2, then yellow in 1*] $= S$[*red in 2*],

(v) And therefore S[*if red in 2, then (if blue in 2, then yellow in 1)*] $= S$.

(vi) Hence, $S \models$ *if red in 2, then (if blue in 2, then yellow in 1)*

Now we have

- $S \models$ *if red in 2, then (if blue in 2, then yellow in 1)*
- $S \models \neg($*if blue in 2, then yellow in 1*$)$
- $S \not\models \neg$*red in 2*

So, *Modus Tollens* fails for conditionals with a conditional as consequent. Modus tollens does not fail if the consequent of the conditional is purely descriptive.

For more counterexamples, see

Yalcin, Seth. (2012). A counterexample to modus tollens. *Journal of Philosophical Logic*, **41**, 1001-1024.

The principle of excluded muddle

Let φ be some descriptive formula, and let S be any state. Then exactly one of the following three possibilities obtains:
(i) $S \models \varphi$
(ii) $S \models \neg\varphi$
(iii) $S \models$ *might* φ, and $S \models$ *might* $\neg\varphi$

This yields nine possible contexts to assert an indicative conditional.

$\varphi \Rightarrow \psi$	ψ	*might* ψ *might* $\neg\psi$	$\neg\psi$
φ	1	2	3
might φ *might* $\neg\varphi$	4	5	6
$\neg\varphi$	7	8	9

Notice that in only in context 5 the acceptability of $\varphi \Rightarrow \psi$ not fixed.

$\varphi \Rightarrow \psi$	ψ	*might* ψ *might* $\neg\psi$	$\neg\psi$
φ	1 ac- cepted	2 re- jected	3 re- jected
might φ *might* $\neg\varphi$	4 ac- cepted	5	6 re- jected
$\neg\varphi$	7 ac- cepted	8 ac- cepted	9 ac- cepted

Some Gricean pragmatics

Maxim of Quality:

don't assert $\varphi \Rightarrow \psi$ in context 2, 3, or 6.

Maxim of Quantity:

don't assert $\varphi \Rightarrow \psi$ in context 1, 4, 7, 8 or 9.

Hence:

context 5 is the normal context for $\varphi \Rightarrow \psi$ to be asserted.

As Strawson said:

"The hypothetical statement carries the implication either of uncertainty about, or disbelief in, the fulfillment of both antecedent and consequent" Strawson(1952)*

Conclusion 1: Pragmatic constraints ensure that an indicative conditional will normally be asserted only if its antecent is relevant to its consequent.

*Strawson, P.F., 1952, *Introduction to Logical Theory*, London, Methuen

Abnormal cases: flouting the maxims

In context 1

If there is one thing I cannot stand, it is getting caught in the rush-hour traffic.

She is fifty if she is a day.

In context 4

This is the best book of the month, if not the year.

There is coffee in the pot if you want some

If there is anything you need, my name is Marcia

In context 8

If it doesn't rain, it pours.

If I don't beat him, I'll trash him.

In context 9

If . . . , I am the emperor of China.

If . . . , I am a Dutchman

If. . . , I'll eat my hat

If. . . , I'll be hanged

Conclusion 2 If you think the paradoxes of implication are real paradoxes, you cannot explain how we ever get to say things like above. The above may be abnormal, degenerate conditionals, they are conditionals nonetheless!

A test for pragmatic incorrectness

Test If the premises of an argument are not compatible with the implicatures of the conclusion, it is pragmatically incorrect.

Compare

premise:	*If there is sugar in the coffee, it will taste fine.*
conclusion:	*If there is sugar and dieseloil in the coffee, it will taste fine.*

premise 1:	*There might be dieseloil in the coffee.*
premise 2:	*If there is sugar in in the coffee, it will taste fine.*
conclusion:	*If there is sugar and dieseloil in the coffee, it will taste fine.*

Never pragmatically correct?

$$\frac{\text{premise 1:}\quad \neg\varphi}{\text{conclusion:}\quad \varphi \Rightarrow \psi}$$

Normally, the conclusion implicates *might* φ, so normally it is pragmatically incorrect. But this is okay:

$$\frac{\text{premise 1:}\quad \neg\varphi}{\text{conclusion:}\quad \varphi \Rightarrow \textit{I'll eat my hat}}$$

Gibbard's Puzzle*

Sly Pete and Mr. Stone are playing poker on a Mississipi riverboat. It is now up to Pete to call or fold. My henchman Zack sees Stone's hand, which is quite good, and signals its content to Pete. My henchman Jack sees both hands, and sees that Pete's hand is rather low, so that Stone's is the winning hand. At this point the room is cleared. A few minutes later, Zack slips me a note which says if Pete called, he won, and Jack slips me a note which says if Pete called, he lost ... I conclude that Pete folded.

* Gibbard, A. (1981). Two recent theories of conditionals. In *Ifs: Conditionals, belief, decision, chance and time*, 211-247, Springer, Dordrecht

For an explanation, see:

Goldstein, Simon. (2022). Sly Pete in Dynamic Semantics',
Journal of Philosophical Logic, **51**, 1103-1117.

Counterfactual conditionals

If it weren't for the war. . .

. . . I almost cheer the day that the Germans invaded Poland, back in 1939. If they hadn't, there would have been no need for my Polish grandfather to go into hiding because he was Jewish. He would not have knocked on the door of my grandmother's house to ask for shelter. They would not have fallen in love, they would not have married, they would not have gotten my father as their son. I would not have existed.*. . .

*Quotation taken (and translated) from 'Dankzij de oorlog', a Dutch play, written by Tomer Pawlicki.

If it weren't for the war...

...I almost cheer the day that the Germans invaded Poland, back in 1939. If they hadn't, there would have been no need for my Polish grandfather to go into hiding because he was Jewish. He would not have knocked on the door of my grandmother's house to ask for shelter. They would not have fallen in love, they would not have married, they would not have gotten my father as their son. I would not have existed.

Compare:

- *If Oswald Kennedy did not kill Kennedy, someone else did.*
- *If Oswald Kennedy had not killed Kennedy, someoneone else would have.*

Compare:

- *If Oswald Kennedy did not kill Kennedy, someone else did*.

- *If Oswald Kennedy had not killed Kennedy, someoneone else would have*.

Counterfactuals

A counterfactual conditional is a sentence of the form

If it had been the case that φ, *it would have been the case that* ψ

Counterfactuals are usually asserted in a context in which the antecedent φ is false, and known to be false by both speaker and addressee.

(So, the riddle is, how is it possible that we can learn something about the actual world by talking about a non-existing situation)

Proverbs

Every language has one or more proverbs expressing the idea that it makes little sense to wonder what would have been the case if. . .

English

If ‘if’s and ‘an’s were pots and pans, there’d be no work for tinkers’ hands.

French

Avec des ‘si’ on mettrait Paris en bouteille
With ‘if’ one could put Paris in a bottle.

German

Wenn das Wörtchen wenn nicht wär, wär mein Vater Millionär.

If this little word ‘if’ had not been there, my father would be a millionaire.

Polish

Gdyby ciocia miała wąsy, byłaby wujaszkiem.

If my aunt had a mustache, she would have been my uncle.

Cantonese

有早知，无乞衣。

(Literary: there is foresight, there is no beggar.)

If one could see into the future, there would be no beggars

Example 1: Counterfactual History

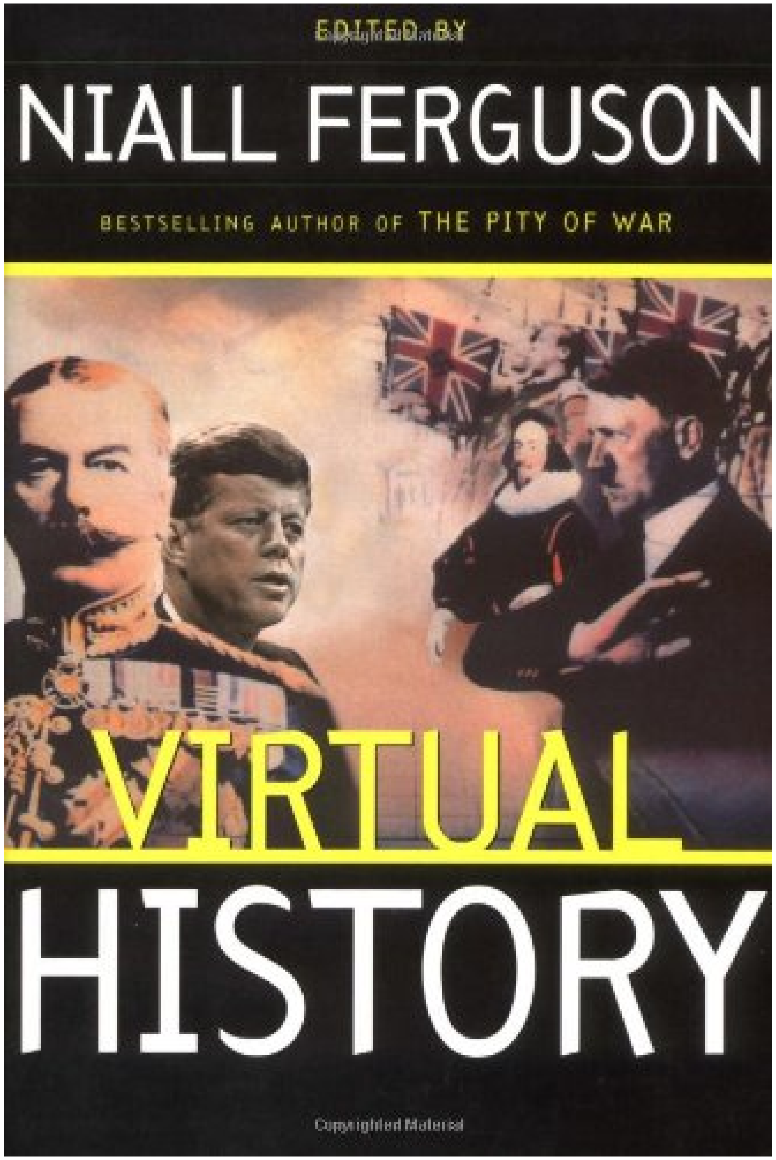
If Deng Xiaoping had not returned to power in 1978, would China's Reform and Opening have unfolded in the same way?

This question helps us see how historical change depends both on structural forces — economic and social pressures — and on individual leadership.

Counterfactual reasoning clarifies why Deng's policies mattered, not by speculation, but by testing the role of contingency in history.

Unscientific!!!

Most historians dislike this kind of statements: they are 'specul-taive', 'unscientific', 'not verifiable' etc.


EDITED BY
NIALL FERGUSON
BESTSELLING AUTHOR OF THE PITY OF WAR
VIRTUAL
HISTORY
Copyrighted Material

The Use of Counterfactual History

Counterfactuals help us

- to uncover causal connections .
- to underscore the role of chance.
- to 'measure' the importance of a given event.

> "It is, at the very root, the idea of conjecturing on what did not happen, or what might have happened, in order to understand what did happen.*

*Jeremy Black; Donald M. MacRaild (2007). Studying History. Palgrave Macmillan. p. 125.

Example 2: Legal accountability

- *If the driver had respected the speed limit, the accident would not have occurred.*

Such a counterfactual is not just ‘hypothetica’l: it is *evidential*. It supports a normative conclusion about liability.

In order to win a lawsuit the plaintiff must prove the legal liability of the defendant. This requires evidence of the duty to act, the failure to fulfill that duty, and the connection of that failure to some injury or harm to the plaintiff.

The defendant’s conduct is a cause of the harm if and only if the harm would not have occurred but for that conduct.

Textbook example

Context*: A man went to a hospital after being poisoned. The hospital negligently failed to examine him, and he later died.

Court's reasoning: The hospital was negligent — but the court held it not liable because, *even if the hospital had treated him properly, the man would still have died from the poison.*

In other words, the judge rejected the counterfactual *If the hospital had treated him, he would have survived.*

The defendant's conduct is only a cause of the harm if the harm would not have occurred but for that conduct.

*Barnett v. Chelsea & Kensington Hospital Management Committee (1969, UK)

Truth condition: what's the problem?

- Counterfactuals cannot be analysed as material implications;

- Counterfactuals cannot be analysed as strict implications.

Logic

From

If there had been sugar in my coffee, it would have tasted better

it does not follow that

If there had been sugar and diesel oil in my coffee, it would have tasted better

(If counterfactuals were strict implications, this would be valid)

Truth condition: basic idea

A sentence of the form

If it had been the case that φ, *it would have been the case that* ψ

is true in world w iff

ψ is true in all possible worlds in which
(a) φ is true, and which
(b) in other respects differ minimally from w.*

*Lewis, David (1973), *Counterfactuals*, Blackwell.

. . . in other respects differ minimally. . .

Tichy's counterexample:*

> 'Consider a man, call him Jones, who is possessed of the following dispositions as regards wearing his hat. Bad weather invariably induces him to wear a hat. Fine weather, on the other hand, affects him neither way: on fine days he puts his hat on or leaves it on the peg, completely at random. Suppose moreover that actually the weather is bad, so Jones *is* wearing his hat.'

The question is: would you accept the sentence *'If the weather had been fine, Jones would have been wearing his hat'*?

*Tichy, P. 'A counterexample to the Stalnaker-Lewis analysis of counterfactuals.' *Philosophical Studies*, 1976.

Reaction

Stalnaker:

> *"...the relevant conception of minimal difference needs to be spelled out with care."*

Lewis:

> *"It is of the first importance to avoid big, widespread, diverse violations of law.*
>
> *It is of little or no importance to secure approximate similarity of particular fact."*

. . . avoid violations of law. . .

Which of the following two sentences is true?

- *If this pen had been made of copper, it would conduct electricity.*
- *If this pen had been made of copper, copper would not conduct electricity.*

avoid violations of ‘law’

‘law’ does not just mean natural law:

- *If White had played 14.Nd5, Black would have lost.*

In making a counterfactual assumption we are not prepared to give up propositions we consider to be general laws. Nor are we prepared to consider situations where chess (or any other game) is not played by the rules, or played by different rules.

. . . particular facts of little or no importance. . .

Suppose that Jones always flips a coin before he opens the curtains to see what the weather is like. Heads means he is going to wear his hat in case the weather is fine, whereas tails means he is not going to wear his hat in that case. Like above, bad weather invariably makes him wear his hat. Now suppose that today heads came up when he flipped the coin, and that it is raining. So, again, Jones is wearing his hat.

Again: Would you accept the sentence *'If the weather had been fine, Jones would have been wearing his hat'*?

So, Lewis was wrong..

Similarity of particular fact *is* important,

but only for facts that do not depend on other facts. Facts stand and fall together. In making a counterfactual assumption, we are prepared to give up everything that depends on something that we must give up to maintain consistency. But we want to keep in as many independent facts as we can.

Ramsey test

This is how to evaluate a counterfactual:

- first, add the antecedent hypothetically to your stock of beliefs;
- second, make whatever adjustments that are required to maintain consistency (without modifying the hypothetical belief in the antecedent);
- finally, consider whether or not the consequent is then true.

15a Note

If two people are arguing if p, will q? & are both in doubt as to p they are ~~but were~~ adding p hypothetically to their stock of knowledge and arguing on that basis about q; so that in a sense if p, q and if p, $\bar{q}$ are contradictories we can say they are fixing their degrees of belief in q given p. If p turns out false these ~~matter~~ degrees of belief are rendered void. If either party believes $\bar{p}$ for certain, the question ceases to mean anything to him except as a question about what follows from certain laws or hypotheses see below pp

Caveat

Assuming *if it had been the case that* φ

$$\neq$$

revising your beliefs by φ

When you believe that φ is true and you try to imagine what would have been the case if φ had been false, you have to change your cognitive state, but it is it not the kind of change you would have to make if you were to discover that φ is *in fact* false. It is not a *correction*.

Example

Consider for example *Oswald killed Kennedy*. Supposing that Oswald had not killed Kennedy might make you think *'If Oswald had not killed Kennedy, Kennedy would have had a second term as president of the US.'*

If, however, at some point you were to find out that your belief that Oswald killed Kennedy is *in fact* wrong, and you had to revise your beliefs accordingly, it is very likely that after this revision you would still believe that Kennedy was killed and therefore did not have a second term as president.

A unified account?

Some philosophers believe that the only difference between indicatives and counterfactuals is that each is used in different circumstances. At the level of semantics, both express the same 'connection' between the antecedent and the consequent.

It would seem that anyone subscribing to this position is committed to the following:

An agent who is ignorant about the truth value of φ, but entitled to entertain the indicative conditional *If* φ, ψ, will later, after learning that φ is in fact false, be entitled to entertain the counterfactual *If it had been the case that* φ*, it would have been the case that* ψ.

Counterexample: From indicative to counterfactual

How can you get from a state in which you believe

- *If the butler did not kill her, somebody else did.*

to a state in which you believe

- *If the butler had not killed her, she would still be alive.*

rather than

- *If the butler had not killed her, somebody else would have done it.*

Update semantics for counterfactuals

Fix a finite set $\mathcal{A}$ of atomic sentences.

- A *world* is a function with domain $\mathcal{A}$ and range $\{0, 1\}$;

- A *situation* is a partial world;

- A *proposition* is a set of worlds;

- If φ is formula of propositional logic, then the proposition expressed by φ is the set $[\![\varphi]\!] = \{w \in \mathcal{W} \mid w(\varphi) = 1\}$.

States

Let $\mathcal{W}$ be the set of possible worlds. A *cognitive state* S is a pair $\langle U_S, F_S \rangle$, where $U_S \subseteq \mathcal{W}$, and $F_S \subseteq U_S$.

- U_S is called the *universe* of the state S. Its extension is determined by the *laws* the agent is acquainted with. F_S represents the agent's *factual information*.

- The *minimal* state, also called the state of ignorance, is given by: $\mathbf{1} = \langle \mathcal{W}, \mathcal{W} \rangle$.

- Whenever $U_S \subseteq U_{S'}$ and $F_S \subseteq F_{S'}$, we say that S *is at least as strong as* S'.

The Minimal State for a language with three atoms

	$p\ q\ r$
w_0	0 0 0
w_1	0 0 1
w_2	0 1 0
w_3	0 1 1
w_4	1 0 0
w_5	1 0 1
w_6	1 1 0
w_7	1 1 1

Interpretation

The interpretation of sentences of propositional logic:

$$S[\varphi] = \langle U_S, F_S \cap [\![\varphi]\!]\rangle$$

The interpretation of laws:

$$S[\Box\varphi] = \langle U_S \cap [\![\varphi]\!], F_S \cap [\![\varphi]\!]\rangle$$

Example

	$p\ q\ r$
w_0	0 0 0
w_1	0 0 1
w_2	0 1 0
w_3	0 1 1
w_4	1 0 0
w_5	1 0 1
w_6	1 1 0
w_7	1 1 1

minimal state

$$\xrightarrow{\substack{\neg q \\ \Box(r \to p)}}$$

	$p\ q\ r$
$\boldsymbol{w_0}$	**0 0 0**
$\cancel{w_1}$	~~0 0 1~~
w_2	0 1 0
$\cancel{w_3}$	~~0 1 1~~
$\boldsymbol{w_4}$	**1 0 0**
$\boldsymbol{w_5}$	**1 0 1**
w_6	1 1 0
w_7	1 1 1

resulting state

Set up

Consider a sentence of the form

if it had been that φ, *it would have been that* ψ

Take a look at the process of interpretation. In interpretating the antecedent in S one gets to S[*if it had been that* φ].
Keep this state in memory, as a subordinate state, so that *it would have been that* ψ or any other sentence in the counter-factual mood can be interpreted in the right context.

More Terminology

Suppose $U \subseteq \mathcal{W}$, and $s \subseteq w$ for some $w \in U$.

- The situation s *determines* the world w *within* U iff for every world v such that $s \subseteq v$ and $v \in U$, it holds that $v = w$;

- The situation s *is a basis for* the world w *within* U iff s is a *minimal* situation determining w within U.

- The situation s *forces* the proposition $[\![\varphi]\!]$ *within* U iff for every world w such that $s \subseteq w$ and $w \in U$, it holds that $w \in [\![\varphi]\!]$.

Retraction 1

- Suppose $w \in [\![\varphi]\!]$. The set $w \downarrow [\![\varphi]\!]$ consists of the situations s with the following property:

  There exists a basis s' for w such that that s is a *maximal* subset of s' not forcing $[\![\varphi]\!]$.

Retraction 2

- $S{\downarrow}[\![\varphi]\!]$, the *retraction of* $[\![\varphi]\!]$ *from* S, is determined as follows:

  (i) $U_{S{\downarrow}[\![\varphi]\!]} = U_S$

  (ii) $v \in F_{S{\downarrow}[\![\varphi]\!]}$ iff $v \in U_S$ and $v \supseteq s$
  for some $w \in F_S$ and some $s \in w \downarrow [\![\varphi]\!]$

- The subordinate state S[*if it had been that* φ], is given by $(S{\downarrow}[\![\neg\varphi]\!])[\varphi]$

Tichy 1

$p :=$ *The weather is bad*; $q :=$ Jones is wearing his hat

The state $S = \mathbf{1}[\Box(p \to q)][p][q]$ can be pictured thus:

	$p\ q\ r$
w_0	0 0 0
w_1	0 0 1
w_2	0 1 0
w_3	0 1 1
$\cancel{w_4}$	$\cancel{1\ 0\ 0}$
$\cancel{w_5}$	$\cancel{1\ 0\ 1}$
$\boldsymbol{w_6}$	**1 1 0**
$\boldsymbol{w_7}$	**1 1 1**

Next we find:

	$p\ q\ r$
w_0	0 0 0
w_1	0 0 1
w_2	0 1 0
w_3	0 1 1
$\cancel{w_4}$	$\cancel{1\ 0\ 0}$
$\cancel{w_5}$	$\cancel{1\ 0\ 1}$
$\boldsymbol{w_6}$	**1 1 0**
$\boldsymbol{w_7}$	**1 1 1**

The basis for w_6 is $\{\langle p, 1\rangle, \langle r, 0\rangle\}$; the basis for w_7 is $\{\langle p, 1\rangle, \langle r, 1\rangle\}$

Furthermore, $w_6 \downarrow [\![p]\!] = \{\{\langle r, 0\rangle\}\}$ and $w_7 \downarrow [\![p]\!] = \{\{\langle r, 1\rangle\}\}$.

This means that $S{\downarrow}[\![p]\!] = \langle U_S, U_S\rangle$.

Therefore, $S[$*if it had been that* $\neg p]$ is given by:

	p q r
$\mathbf{w_0}$	$\mathbf{0\ 0\ 0}$
$\mathbf{w_1}$	$\mathbf{0\ 0\ 1}$
$\mathbf{w_2}$	$\mathbf{0\ 1\ 0}$
$\mathbf{w_3}$	$\mathbf{0\ 1\ 1}$
$\cancel{w_4}$	$\cancel{1}\ \cancel{0}\ \cancel{0}$
$\cancel{w_5}$	$\cancel{1}\ \cancel{0}\ \cancel{1}$
w_6	$1\ 1\ 0$
w_7	$1\ 1\ 1$

	p q r
$\boldsymbol{w_0}$	**0 0 0**
$\boldsymbol{w_1}$	**0 0 1**
$\boldsymbol{w_2}$	**0 1 0**
$\boldsymbol{w_3}$	**0 1 1**
$\cancel{w_4}$	$\cancel{1\ 0\ 0}$
$\cancel{w_5}$	$\cancel{1\ 0\ 1}$
w_6	1 1 0
w_7	1 1 1

Notice that S[*if it had been that* $\neg p$] $\not\models q$.

Therefore,
$S \not\models$ *if it had been that* $\neg p$, *it would have been that* q

Tichy 2

$p :=$ *The weather is bad*
$q :=$ *Jones is wearing his hat*
$r :=$ *The coin comes up heads*

The state $S = \mathbf{1}[\Box((p \vee r) \leftrightarrow q)][r][p]$ is given by:

	$p\ q\ r$
w_0	0 0 0
$\cancel{w_1}$	$\cancel{0}\ \cancel{0}\ \cancel{1}$
$\cancel{w_2}$	$\cancel{0}\ \cancel{1}\ \cancel{0}$
w_3	0 1 1
$\cancel{w_4}$	$\cancel{1}\ \cancel{0}\ \cancel{0}$
$\cancel{w_5}$	$\cancel{1}\ \cancel{0}\ \cancel{1}$
w_6	1 1 0
$\boldsymbol{w_7}$	**1 1 1**

	$p\ q\ r$
w_0	0 0 0
$\cancel{w_1}$	$\cancel{0}\ \cancel{0}\ \cancel{1}$
$\cancel{w_2}$	$\cancel{0}\ \cancel{1}\ \cancel{0}$
w_3	0 1 1
$\cancel{w_4}$	$\cancel{1}\ \cancel{0}\ \cancel{0}$
$\cancel{w_5}$	$\cancel{1}\ \cancel{0}\ \cancel{1}$
w_6	1 1 0
$\boldsymbol{w_7}$	**1 1 1**

$\xrightarrow{\textit{if had been } \neg p}$

	$p\ q\ r$
w_0	0 0 0
$\cancel{w_1}$	$\cancel{0}\ \cancel{0}\ \cancel{1}$
$\cancel{w_2}$	$\cancel{0}\ \cancel{1}\ \cancel{0}$
$\boldsymbol{w_3}$	**0 1 1**
$\cancel{w_4}$	$\cancel{1}\ \cancel{0}\ \cancel{0}$
$\cancel{w_5}$	$\cancel{1}\ \cancel{0}\ \cancel{1}$
w_6	1 1 0
w_7	1 1 1

The basis for w_7 is $\{\langle p, 1\rangle, \langle r, 1\rangle\}$
Furthermore, $w_7 \downarrow [\![p]\!] = \{\{\langle r, 1\rangle\}\}$.
This means that $S \downarrow [\![p]\!] = \langle U_S, \{w_3, w_7\}\rangle$.

Therefore, the subordinate state $S[$*if it had been that* $\neg p]$
is given by the table on the right.

	$p\ q\ r$
w_0	0 0 0
$\cancel{w_1}$	$\cancel{0}\ \cancel{0}\ \cancel{1}$
$\cancel{w_2}$	$\cancel{0}\ \cancel{1}\ \cancel{0}$
$\boldsymbol{w_3}$	$\mathbf{0\ 1\ 1}$
$\cancel{w_4}$	$\cancel{1}\ \cancel{0}\ \cancel{0}$
$\cancel{w_5}$	$\cancel{1}\ \cancel{0}\ \cancel{1}$
w_6	1 1 0
w_7	1 1 1

Notice that S[*if it had been that* $\neg p$] $\models q$.

Therefore,
$S \models$ *if it had been that* $\neg p$, *it would have been that* q

The butler 1

p := *The butler killed the duchess* q := *The gardener killed the duchess* r := *The duchess was killed*

Consider first the state $S = \mathbf{1}[\Box((p \vee q) \to r)][r][p \vee q]$.

	$p\ q\ r$
w_0	0 0 0
w_1	0 0 1
$\cancel{w_2}$	$\cancel{0}\ \cancel{1}\ \cancel{0}$
$\mathbf{w_3}$	**0 1 1**
$\cancel{w_4}$	$\cancel{1}\ \cancel{0}\ \cancel{0}$
$\mathbf{w_5}$	**1 0 1**
$\cancel{w_6}$	$\cancel{1}\ \cancel{1}\ \cancel{0}$
$\mathbf{w_7}$	**1 1 1**

	p q r
w_0	0 0 0
w_1	0 0 1
$\cancel{w_2}$	$\cancel{0}$ $\cancel{1}$ $\cancel{0}$
$\mathbf{w_3}$	**0 1 1**
$\cancel{w_4}$	$\cancel{1}$ $\cancel{0}$ $\cancel{0}$
$\mathbf{w_5}$	**1 0 1**
$\cancel{w_6}$	$\cancel{1}$ $\cancel{1}$ $\cancel{0}$
$\mathbf{w_7}$	**1 1 1**

Notice that $S[\neg p] \models q$.
So, on our account of indicative conditionals, $S \models$ *If* $\neg p, q$.

The butler 2

Next, consider the state $S' = \mathbf{1}[\Box((p \vee q) \to r)][r][p \vee q][p][\neg q]$, pictured below on the left hand side.

	$p\ q\ r$
w_0	0 0 0
w_1	0 0 1
$\cancel{w_2}$	$\cancel{0}\ \cancel{1}\ \cancel{0}$
w_3	0 1 1
$\cancel{w_4}$	$\cancel{1}\ \cancel{0}\ \cancel{0}$
$\boldsymbol{w_5}$	$\mathbf{1\ 0\ 1}$
$\cancel{w_6}$	$\cancel{1}\ \cancel{1}\ \cancel{0}$
w_7	1 1 1

$\xrightarrow{\textit{if had been } \neg p}$

	$p\ q\ r$
$\boldsymbol{w_0}$	$\mathbf{0\ 0\ 0}$
$\boldsymbol{w_1}$	$\mathbf{0\ 0\ 1}$
$\cancel{w_2}$	$\cancel{0}\ \cancel{1}\ \cancel{0}$
w_3	0 1 1
$\cancel{w_4}$	$\cancel{1}\ \cancel{0}\ \cancel{0}$
w_5	1 0 1
$\cancel{w_6}$	$\cancel{1}\ \cancel{1}\ \cancel{0}$
w_7	1 1 1

	$p\ q\ r$
w_0	0 0 0
w_1	0 0 1
$\cancel{w_2}$	~~0 1 0~~
w_3	0 1 1
$\cancel{w_4}$	~~1 0 0~~
$\boldsymbol{w_5}$	**1 0 1**
$\cancel{w_6}$	~~1 1 0~~
w_7	1 1 1

$$\xrightarrow{\textit{if had been } \neg p}$$

	$p\ q\ r$
$\boldsymbol{w_0}$	**0 0 0**
$\boldsymbol{w_1}$	**0 0 1**
$\cancel{w_2}$	~~0 1 0~~
w_3	0 1 1
$\cancel{w_4}$	~~1 0 0~~
w_5	1 0 1
$\cancel{w_6}$	~~1 1 0~~
w_7	1 1 1

The state $S'[\textit{if had been that } \neg p] \not\models q$.
In other words, $S \not\models$ *if it been that* $\neg p$, *would have been that* q

The above illustrates that there is a huge difference between making a counterfactual assumption and correcting one's beliefs.

A problem

The circuit example Suppose there is a circuit such that the light is on (r) exactly when both switches are in the same position (up or not up). At the moment switch 1 is down ($\neg p$), switch 2 is up (q) and the lamp is off ($\neg r$).

Now, intuitively, *if switch 1 had been up, the light would have been on*, right?

Lifschitz, V., 'Frames in the space of situations', *Artificial Intelligence* 46: 365-376, 1986

In a picture

	$p\ q\ r$
$\cancel{w_0}$	$\cancel{0}\ \cancel{0}\ \cancel{0}$
w_1	0 0 1
$\boldsymbol{w_2}$	**0 1 0**
$\cancel{w_3}$	$\cancel{0}\ \cancel{1}\ \cancel{1}$
w_4	1 0 0
$\cancel{w_5}$	$\cancel{1}\ \cancel{0}\ \cancel{1}$
$\cancel{w_6}$	$\cancel{1}\ \cancel{1}\ \cancel{0}$
w_7	1 1 1

Unfortunately, as you can check,on our account
$S \not\models$ *if had been* p, *would have been* r
What we get is $S \models$ *if had been* p, *might have been* $(q \wedge r)$,
and $S \models$ *if had been* p, *might have been* $(\neg q \wedge \neg r)$

Compare

Three sisters Consider the case of the three sisters who live together, and basically all of the time are home together, except when every now and then two of them go out for shopping. In those cases it can be any two of them who go out, and the third one stays home.

At the moment Ronny is out ($\neg r$) together with Paula ($\neg p$) while Quinzy is at home (q).

Suppose now counterfactually that Paula had been at home. . .

Would you in this case say: *Well, in that case Ronny would have been at home, too*? I guess not. After all, she might have gone shopping with Quinzy instead of Paula.

Something is missing

Formally, the three sisters case and the case of the switches are equal. But intuitively, there is a difference. This means that our set up is still too crude.

Katrin Schulz* and Stefan Kaufmann† think that in the switch example causality plays a crucial role.

*Schulz, Katrin. (2011). If you'd wiggled A, then B would've changed. Causality and counterfactual conditionals, *Synthese* **179**, 239-251.

†Kaufmann, Stefan. (2013). Causal premise semantics. *Cognitive science* **37**, 1136-1170.