UNIVERSITY OF AMSTERDAM

Nixon, Light Switches and King Ludwig of Bavaria: How to Model Counterfactual Reasoning

Katrin Schulz

1 Conditionals between disciplines

Conditionals

Linguistics

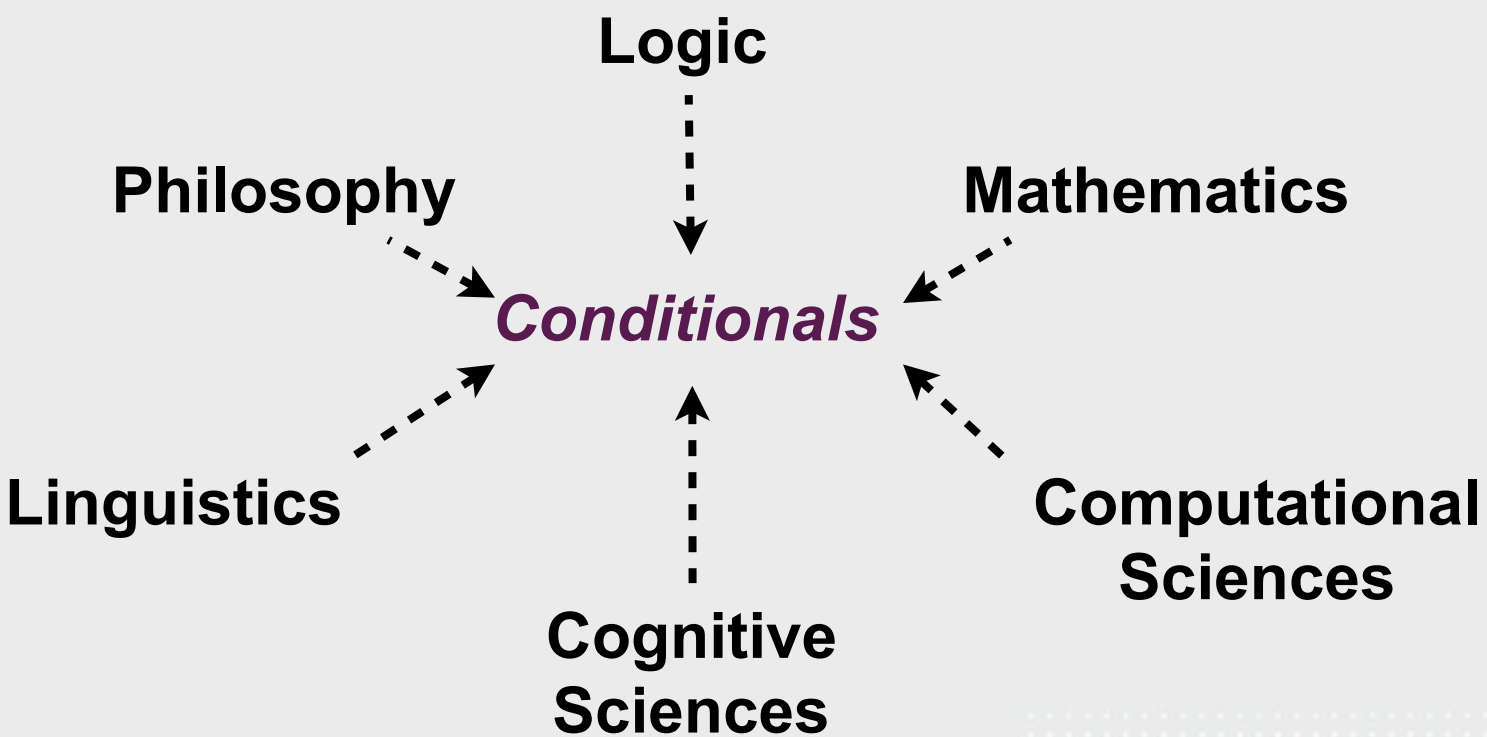
1 Conditionals between disciplines

"I will be discussing a kind of conditional ... typically expressed in English by subjunctive conditionals. Here are some examples: 'if I were to strike this match there would be an explosion', ... This kind of counterfactual is intimately connected with **laws, explanation, causation, choice, knowledge, memory, measurement, chance, the asymmetry of past and future**, etc; a veritable Who's Who of philosophically and scientifically significant concepts. Philosophers may disagree about the order of explanation among these items and counterfactuals but everyone ought to agree that we would make significant progress understanding them all if we had an account of what makes this kind of counterfactual statement true/false." (B. Loewer)

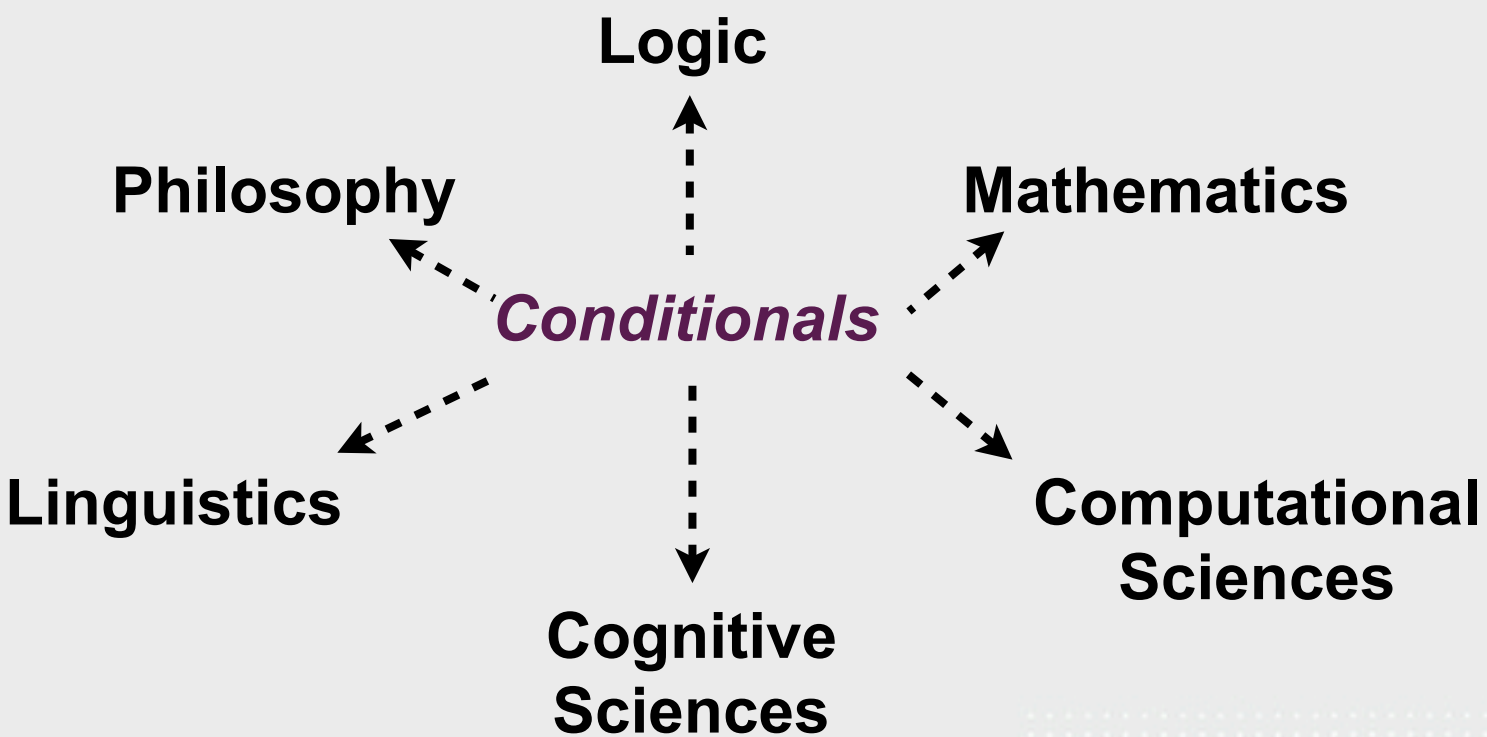
1 Conditionals between disciplines

- Conditionals give ***concrete form*** to ***abstract reasoning.***
- They are basically everywhere.

1 Conditionals between disciplines



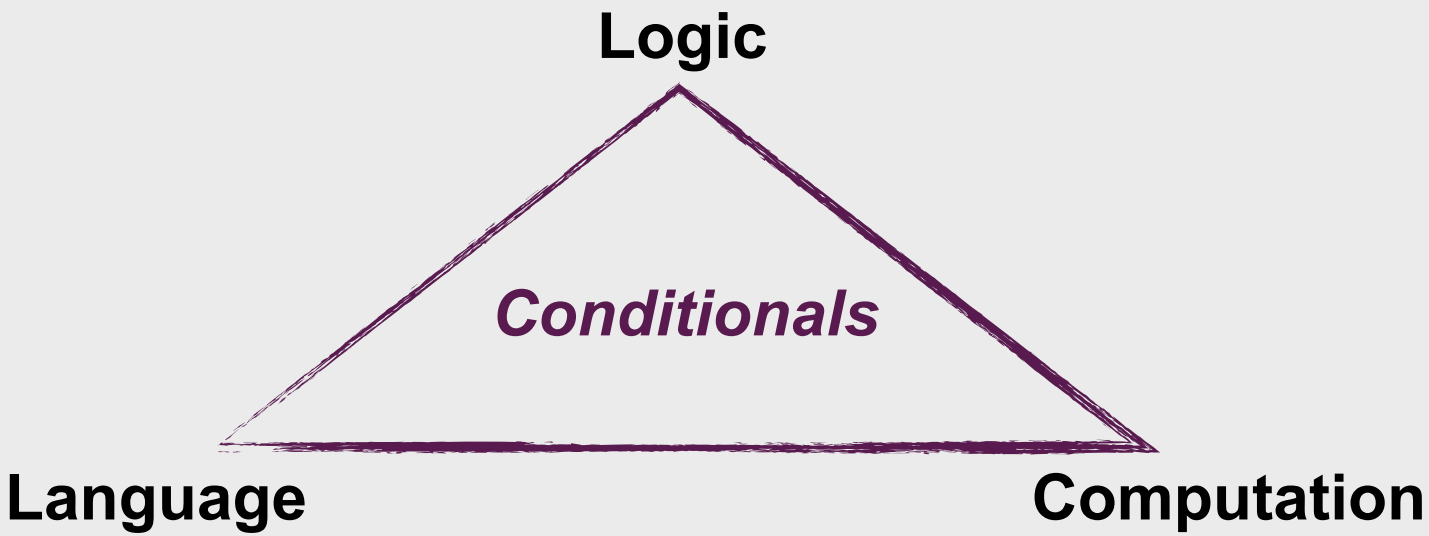
1 Conditionals between disciplines



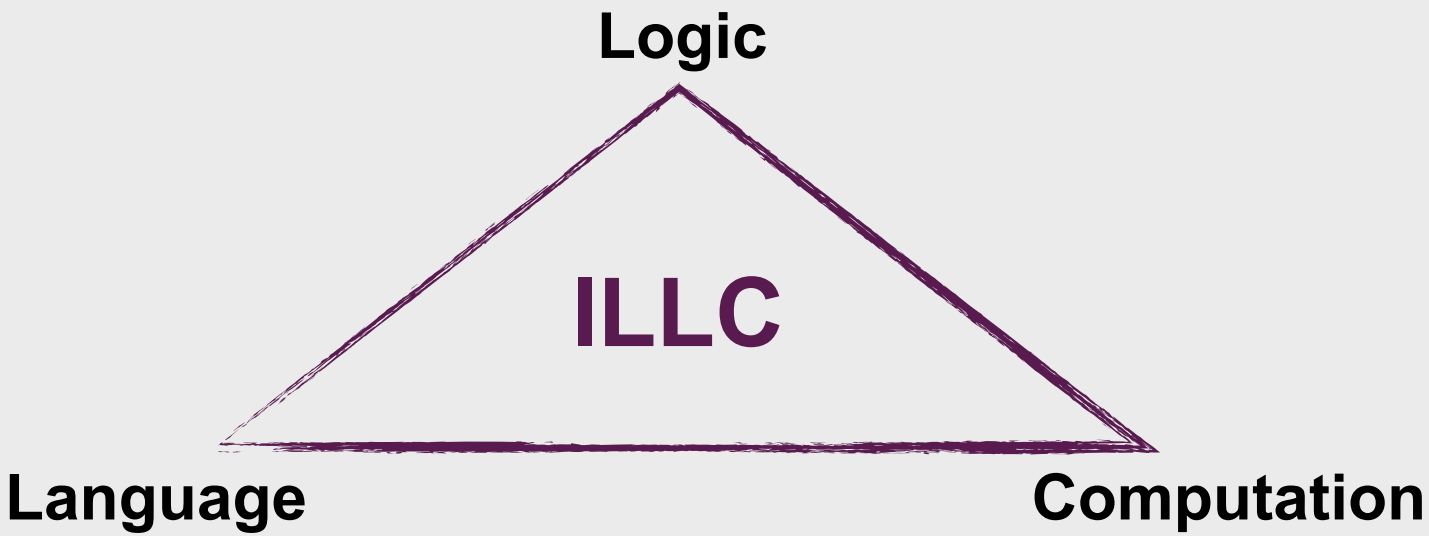
1 Conditionals between disciplines

- Every application improves the theory !

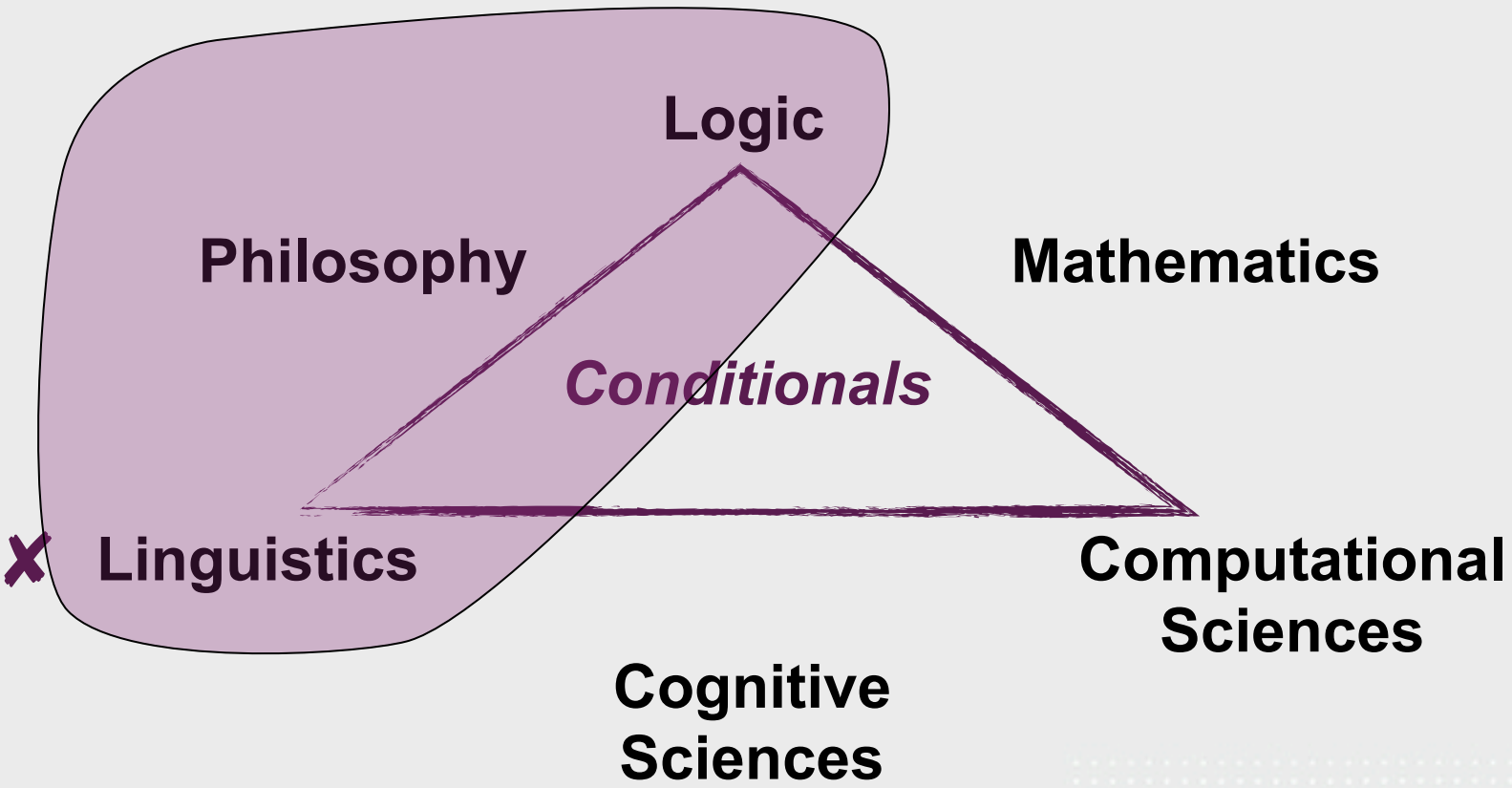
1 Conditionals between disciplines



1 Conditionals between disciplines



1 Conditionals between disciplines



1 Conditionals between disciplines

Goals today

- give you an impression of the fascination conditionals exert on scholars of various disciplines
- explain to you the generals ideas of and motivation behind a number of approaches to the meaning of conditionals

A semantic problem ...

... and its philosophical analysis

2.1 The semantic problem

Goal: Give a formally precies description of the meaning of counterfactual conditionals.

2.2 What is a counterfactual?

"Counterfactual conditionals are sentences of the form
If it had been the case that A; it would have been the case that C.
They are typically uttered in contexts where the antecedent is false and known to be false." [Veltman]

➡ hybrid definition: form and meaning

(1) If I were you, I wouldn't do that.

(2) If she had taken Arsenic, she would have shown exactly the symptoms she is showing.

2.2 What is a counterfactual?

- *We will study:* counterfactual conditionals, i.e. conditionals with a false antecedent.

➡ counterfactuals are fascination because they talk about something that *is not*

2.1 The semantic problem

Goal: Give a formally precies description of the meaning of counterfactual conditionals.

A counterfactual conditional '*If A had been the case, then C would have been the case*' is true given model M and world w_0 iff: ...

$M, w_0 \models A \mathrel{\leadsto} C$ *iff* ...

➡ huge linguistic problem: what is the relation to natural language counterfactuals

2.1 The semantic problem

Goal: Give a formally precies description of the meaning of counterfactual conditionals.

A counterfactual conditional '*If A had been the case, then C would have been the case*' is true given model M and world w_0 iff: ...

$M, w_0 \models A \mathrel{>\!\!\!\sim} C$ *iff* ...

Assumption: 1. truth-conditional semantics

2. possible world semantics

2.1 The semantic problem

Consider the following case. We think that the local zoo might get a new animal this spring, but have different hunches about what it would be. I suspect an armadillo and you a roadrunner. We like to bet so I wager $5 that

(29) If the zoo gets an animal this spring it will be an armadillo.

You wager against me. Spring comes. It brings wild flowers, birds, bees, but no new animal to the zoo. Who has to pay? The intuition that neither of us gets paid is overwhelming. This remains even if we find out that the zoo board had decided to get an armadillo but the funding was cut at the last minute. I made the better bet, but my attempts to collect $5 may be rebuffed.

2.1 The semantic problem

Goal: Give a formally precies description of the meaning of counterfactual conditionals.

A counterfactual conditional '*If A had been the case, then C would have been the case*' is true given model M and world w_0 iff: ...

$M, w_0 \models A \mathrel{\leadsto} C$ *iff* ...

➡ Be aware of the decisions your are making just by following the obvious!

2.3 What clearly doesn't work

Material Implication

$M, w_o \models A \rightarrow C$ *iff* $M, w_o \not\models A$ or $M, w_o \models C$

	[[A]]	[[C]]
w1	0	0
w2	0	1
w3	1	0
w4	1	1

Problems:

- no relation between antecedent and consequent
- counterfactuals are trivially true

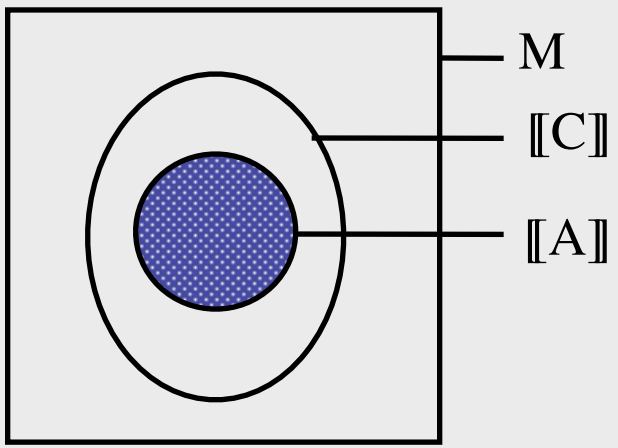
(3) If Amsterdam was a river, I would be wearing a blue hat today.

(4) If I had dropped this glass, it would have grown wings and flown away.

2.3 What clearly doesn't work

Strict conditional

$M, w_0 \models A \succ C$ *iff* $[\![A]\!]^M \subseteq [\![C]\!]^M$



Problems:

- independent of evaluation world
- too strong; not all A-worlds should be considered

(3) If that match had been scratched, it would have lighted.

➔ *doesn't consider worlds where the match is wet or broken.*

2.4 Logical properties of counterfactuals

Strengthening of A: If $A \succ C$, then $(A \wedge B) \succ C$

Contraposition: If $A \succ C$, then $\neg C \succ \neg A$

Transitivity: If $A \succ B$ and If $B \succ C$, then $A \succ C$

(4) If this match were struck, it would light, but if this match had been soaked in water overnight and it were struck, it wouldnt light.

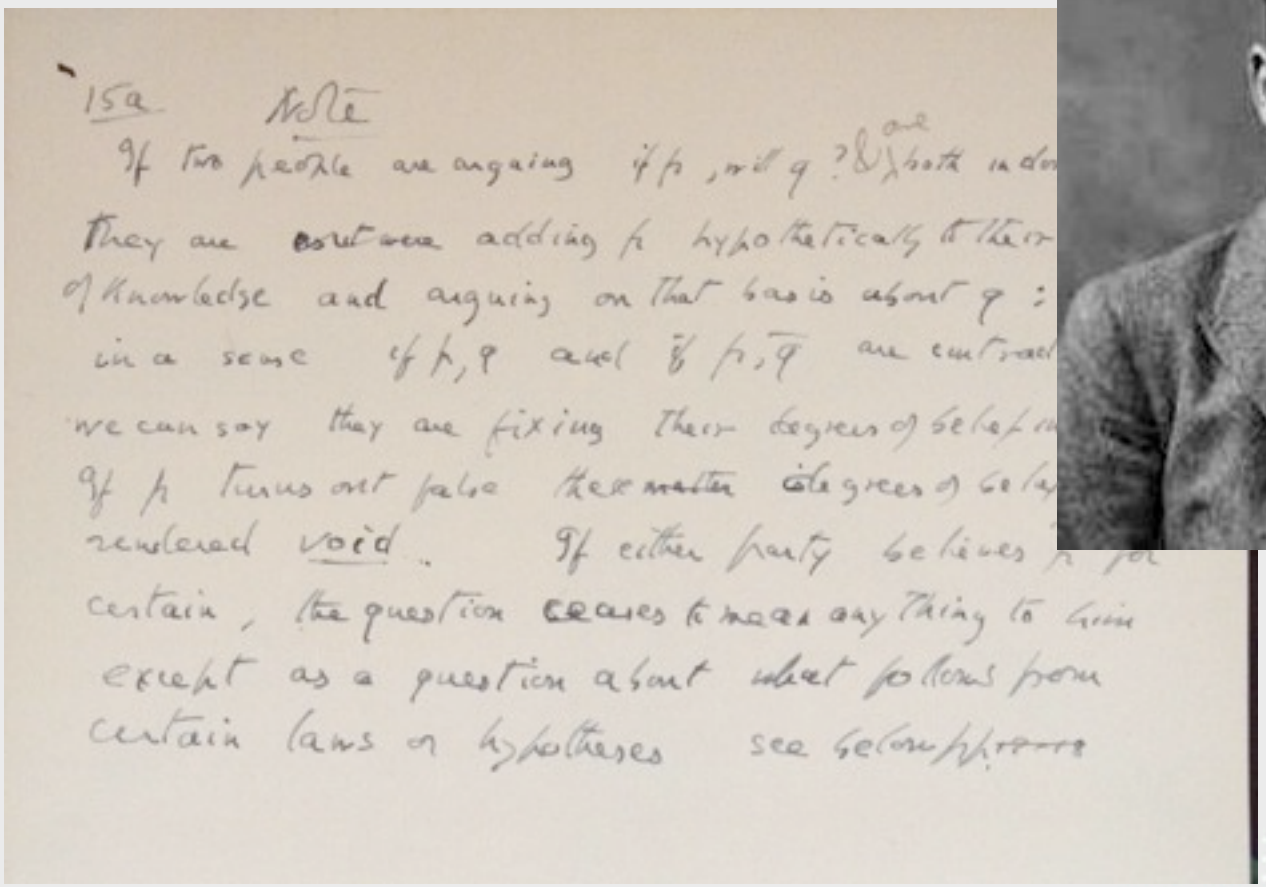
(5) (Even) if Goethe had survived the year 1832, he would be dead by now. ⇏ If Goethe were not dead by now, he would not have survived the year 1832.

(6) If Jones wins the election, Smith will retire to private life. If Smith dies tomorrow, Jones will win the election. ⇏ If Smith dies tomorrow, Smith will retire to private life.

A semantic problem ...

... and its philosophical analysis

3 The similarity approach


15a Note

If two people are arguing if p, will q? & both in ador

They are ~~sometimes~~ adding p hypothetically to their

of knowledge and arguing on that basis about q:

in a sense if p, q and if p, ~q are contrad

we can say they are fixing their degrees of belief in

If p turns out false ~~their~~ their degrees of belief

rendered void. If either party believes p for

certain, the question ~~ceases~~ ceases to mean anything to him

except as a question about what follows from

certain laws or hypotheses see below/overleaf

3 The similarity approach

Ramsey receipt

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.*

3 The similarity approach

hypothetical
change of beliefs

epistemic vs. **ontic** reading of Ramsey's receipt

hypothetical
change of facts

3 The similarity approach

"The duchess has been murdered, and you are supposed to find the murderer. At some point only the butler and the gardener are left as suspects. At this point you believe

(1) If the butler did not kill her, the gardener did.

Still, somewhat later – after you found out convincing evidence showing that the butler did it, and that the gardener had nothing to do with it – you get in a state, in which you will reject the sentence

(2) If the butler had not killed her, the gardener would have."

3 The similarity approach

"Suppose that one Sunday night you approach a small town of which you know that it has exactly two snackbars. Just before entering the town you meet a man eating a hamburger. You have good reason to accept the following indicative conditional:

(7) *If snackbar A is closed, then snackbar B is open.*

Suppose now that after entering the town, you see that A is in fact open. Would you now accept the following conditional?

(8) *If snackbar A were closed, then snackbar B would be open."*

3 The similarity approach

- empirical situation unclear:

only ontic readings ⟷ *only epistemic readings*

- clear consequences for the logic

local revision ⟷ *global revision*

3 The similarity approach

- we focus on the ontic reading of counterfactual conditionals
- we will discuss the ontic formlization of the Ramsey receipt
- this is the Stalnaker/Lewis similarity approach to counterfactuals

3 The similarity approach

Basic Idea

A sentence

> *If it had been the case that A; it would have been the case that C*

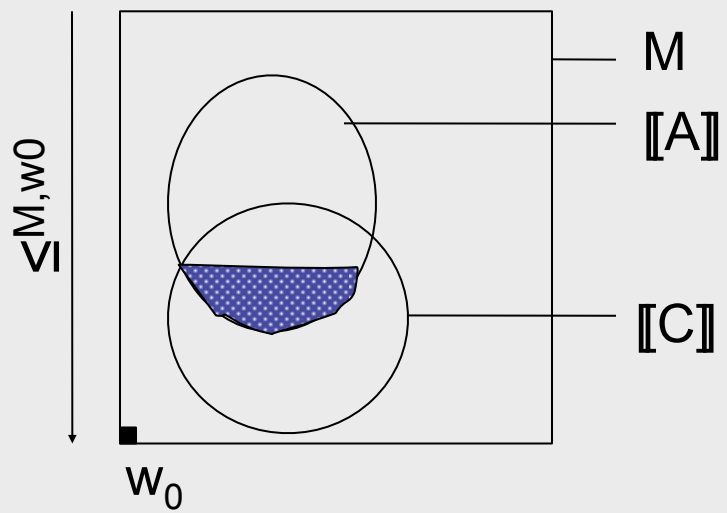
is true in the actual world w_0 iff C is true in all possible worlds in which

(a) A is true, and which

(b) in other respects are maximally similar to w_0.

3 The similarity approach - Problems

$$M, w_0 \models A \succ C \text{ iff } Min(\leq^{M,w_0}, [\![A]\!]^M) \subseteq [\![C]\!]^M$$



Problem:

- How to define the order $\leq^{M,w0}$?

3 The similarity approach - Problems

"The counterfactual

(1) *If Nixon had pressed the button there would have been a nuclear holocaust.*

is true or can be imagined to be so. Now suppose that there never will be a nuclear holocaust. Then that counterfactual is, on Lewis' analysis, very likely false. For given any world in which the antecedent and consequent are both true it will be easy to imagine a closer world in which the antecedent is true and the consequent false. For we need only imagine a change that prevents the holocaust but that does not require such a great divergence from reality." (Fine 1975: 452)

3 The similarity approach - Problems

"Consider a man - call him Jones who is possessed of the following disposition as regards wearing his hat. If the man on the news predicts bad weather, Mr Jones invariably wears his hat the next day. A weather forecast in favor of fine weather, on the other hand, affects him neither way: in this case he puts his hat on or leaves it on the peg, completely at random.

Suppose, moreover, that yesterday bad weather was prognosed, so Jones is wearing his hat. In this case, ...

(1) *If the weather forecast had been in favor of fine weather, Jones would have been wearing his hat*." [Tichy (1976)]

3 The similarity approach - Problems

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."

3 The similarity approach - Problems

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 before, 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.

Would you accept the statement:

(4) If the weather had been fine, Jones would have been wearing his hat.

A semantic problem ...

... and an even better answer

4 Premise Semantics (Goodman, Veltman, Kratzer)

- similarity is defined in terms of a selected set of singular facts of the evaluation world
  - ➡ the premise function P
- in moving to a counterfactual scenario we maintain the general laws of the evaluation world

Similarity relation:

$$w_1 \leq^{M,wo} w_2 \text{ iff } \{\phi \in P(w_o) : M, w_1 \models \phi\} \supseteq \{\phi \in P(w_o) : M, w_2 \models \phi\}$$

➡ special case of the similarity approach

4 Premise Semantics (Goodman, Veltman, Kratzer)

$M, w_o \models A \mathrel{\sim\!\!>} C$ iff $Rev(P(w_o), A) \cup L \models C$

1. take maximal subsets of $P(w_o)$ consistent with A
2. check whether this together with A and L entails C.

Problems:

- what are the laws L?
- what are the premises $P(w_0)$?

☑ The premises can't be all true statements about w_0 (Veltman '76).

4 Premise Semantics (Veltman 2005)

- The relevant set of singular facts is given by revising the BASIS $B_L(w_0)$ of the evaluation world with the antecedent ($Rev_L(B, A)$)

> Basis(w_o) = minimal set of primitive facts of w_o from which, given L, all other facts of w_o can be derived.

- Particular variant of Premise Semantics

4 Premise Semantics (Veltman 2005)

Tichy's Mr. Jones example

Mr. Jones is possessed of the following disposition as regards wearing his hat. If the man in the news predicts bad weather, Mr. Jones invariably wears his hat the next day. If the weather forecast is in favor of fine weather, he puts his hat on or leaves it on the peg completely at random. Suppose, moreover, that yesterday bad weather was prognosed, so Jones is wearing his hat. In this case ...

(2) If the weather forecast had been in favor of fine weather, Jones would have been wearing his hat.

4 Premise Semantics (Veltman 2005)

Tichy's Mr. Jones example

Rule: *bad* → *hat*
world w_0: *bad, hat*
Conditional: ¬ *bad* > *hat*

$B_L(w_0)$ = { bad } ⟹ Basis of w_0

$Rev_L(B, \neg\ bad\ )$ = ∅ ⟹ relevant facts of w_0

Antecedent	+	$Rev_L(B_L(w_0),A)$	$\Rightarrow_L$	Consequent

4 Premise Semantics (Veltman 2005)

Tichy's Mr. Jones example

Rule: *bad → hat*
world w_0: *bad, hat*
Conditional: ¬ *bad* > *hat* ✔

$B_L(w_0)$ = { bad } → Basis of w_0

$Rev_L(B, \neg\ bad\ )$ = ∅ → relevant facts of w_0

4 Premise Semantics (Veltman 2005)

- Veltman 2005 correctly predicts both variants of the Tichy example
- The approach also correctly predicts the Nixon example, and that in the duchess example the counterfactual comes out as false.
- However, also this approach has to face some challenges ...

4 Premise Semantics (Veltman 2005)

Back to Lifschitz's Circuit Example

Suppose there is a circuit such that the light is on (L) exactly when both switches are in the same position (up or not up). At the moment switch 1 is down (¬S1), switch two is up (S2) and the lamp is out (L).

(1) If switch 1 had been up, the lamp would have been on.

4 Premise Semantics (Veltman 2005)

Back to Lifschitz's Circuit Example

Rule: $(S1 \leftrightarrow S2) \leftrightarrow L$
world w_0*:* $\neg S1, S2, \neg L$
Conditional: $S1 > L$

$B_R(w_0) =$ $\{\{\neg S1, S2\}, \{\neg S1, \neg L\}, \{S2, \neg L\}\}$ → B_1, B_2, B_3

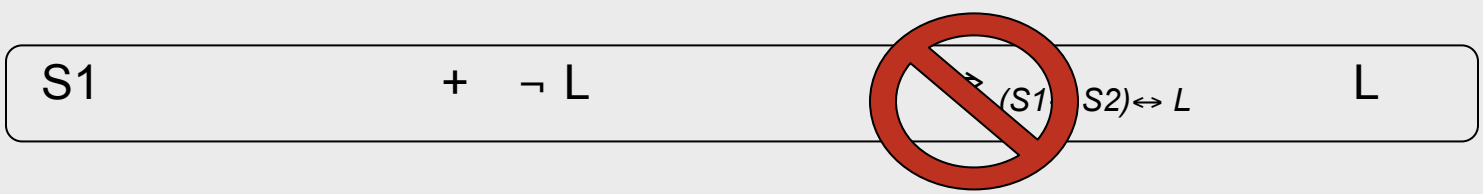
$Rev_R(B_1, S1) = \{S2\}$
$Rev_R(B_2, S1) = \{\neg L\}$
$Rev_R(B_3, S1) = \{S2\}, \{\neg L\}$

→ relevant facts of w_0

4 Premise Semantics (Veltman 2005)

Back to Lifschitz's Circuit Example

Rule: $(S1 \leftrightarrow S2) \leftrightarrow L$
world w_0: $\neg S1, S2, \neg L$
Conditional: $S1 > L$ ✘



$Rev_R(B_1, S1) = \{S2\}$ ✔
$Rev_R(B_2, S1) = \{\neg L\}$ ✘
$Rev_R(B_3, S1) = \{S2\}$ ✔, $\{\neg L\}$ ✘

→ relevant facts of w_0

4 Premise Semantics (Veltman 2005)

Back to Lifschitz's Circuit Example

Rule: $(S1 \leftrightarrow S2) \leftrightarrow L$
world w_0: $\neg S1, S2, \neg L$
Conditional: $S1 > L$

$BASIS_R(w_0) = \{ \{\neg S1, S2\}, \{\neg S1, \neg L\}, \{S2, \neg L\}\}$

B_1
B_2
B_3

Basis
causally indep. facts

$Rev_R(B_1, S1) = \{S2\}$
$Rev_R(B_2, S1) = \{\neg L\}$
$Rev_R(B_3, S1) = \{S2\}, \{\neg L\}$

relevant facts
of w_0

4 Premise Semantics (Veltman 2005)

Solution

- The basis B of the evaluation world w_0 is the set of primitive facts (literals) of w) from which, given the rules R, all other facts of w0 can be derived.

Not: epistemic derivation
But: causal derivation

5 Causal Premise Semantics (Schulz 2007, 2011)

Central Claim

The semantics of (the ontic reading of) counterfactuals relies on a CAUSAL notion of consequence.

UNIVERSITY OF AMSTERDAM

The semantics of counterfactuals ...

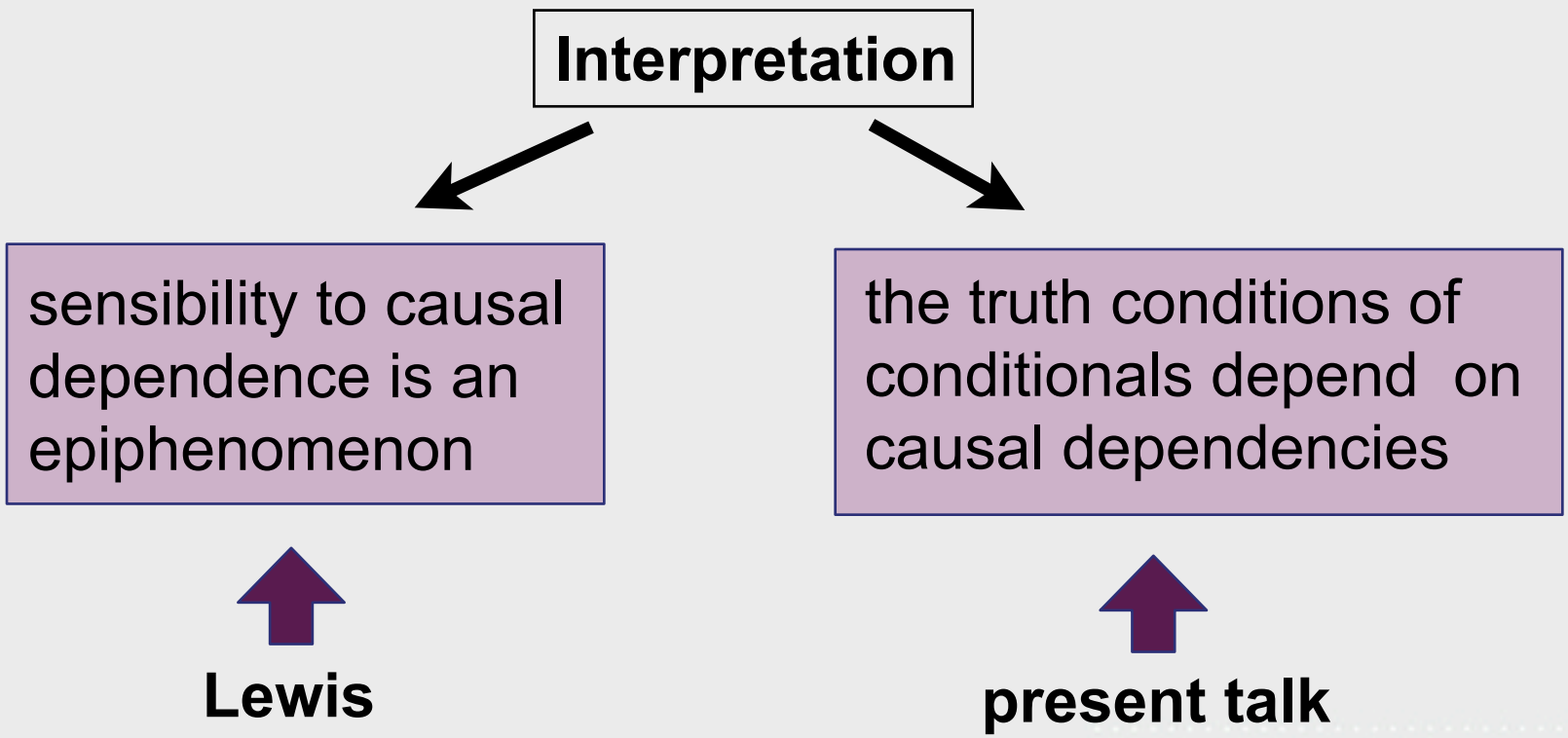
... a causal approach

5 Causal Premise Semantics (Schulz 2007, 2011)

Observation

Conditionals reason along causal dependencies.

5 Causal Premise Semantics (Schulz 2007, 2011)



3 Causal reasoning
3.1 Introduction

NEED:

- a richer notion of model that contains a representation D of direct causal dependencies,

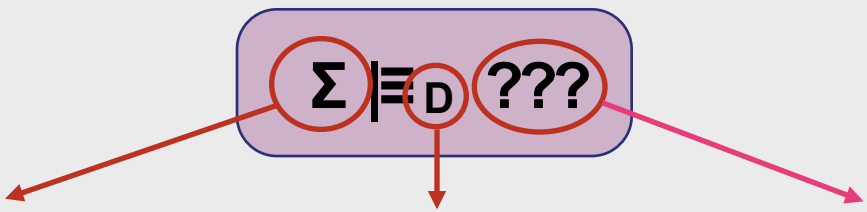
← causal networks, Pearl '00

- a causal notion of consequence $|\equiv$.

← logic programming, van Lambalgen et al.

3 Causal reasoning
3.2 A causal notion of entailment

GOAL: define a causal notion of entailment



primitive facts

model of causal dependencies

causally entailed facts

4 The semantics of conditionals
4.1 The big picture

A conditional sentence '*If A then C*' is true iff:

Antecedent + Facts of w_0 $\Rightarrow_L$ Consequent

basis of w_0

causal entailment

$A \quad B_D(w_0) \quad |\equiv_D \quad C$

4 The semantics of conditionals
4.2 The basis

Definition: ***basis***

The basis $B_L(w_0)$ of the evaluation world w_0 is the minimal set of primitive facts (literals) of w_0 from which all other facts of w_0 follow.

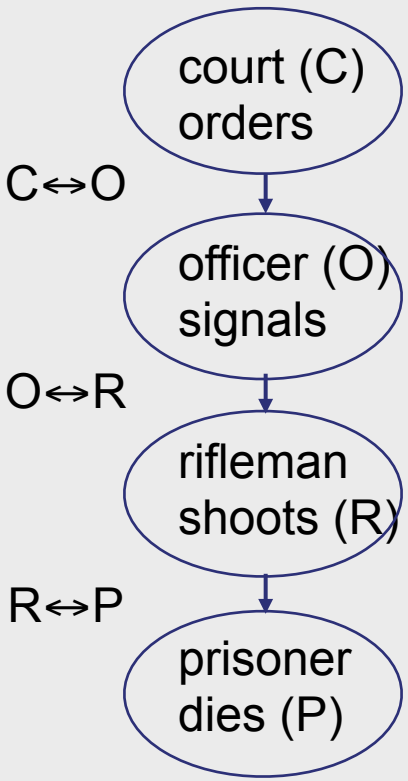
Veltman '05

causally follows.

4 The semantics of conditionals
4.2 The basis

There is a court, an officer, a rifleman and a prisoner. If the court orders the execution of the prisoner, the officer will give a signal to the rifleman, the rifleman will shoot and the prisoner will die.

4 The semantics of conditionals
4.2 The basis



	C	O	R	P
w0	(1)	1	1	1
w1	(1)	(0)	0	0
w2	(1)	1	(0)	(1)

4 The semantics of conditionals
4.3 The big picture again

A conditional sentence '*If A then C*' is true iff:

Antecedent + Facts of w_0 $\Rightarrow_L$ Consequent

$A \quad \cup C \quad B_D(w_0) \quad |\equiv_D \quad C$

4 The semantics of conditionals
4.5 Summary

A conditional sentence '*If A then C*' is true iff:

$$A \ {}^{\cup}C \quad B_D(w_0) \ |\equiv_D \ C$$

- ***Causal Premise Semantics***
  - *No matter how you force A in w_0 (by intervention), C will causally follow.*

UNIVERSITY OF AMSTERDAM

Philosophical Assessment

5 Philosophical Assessment
5.1 Predictions

- Approach makes the correct predictions for the target examples
  - backtracking
  - complex dependencies (circuit example)
- Approach can account for critical data beyond the primary target
  - Kit Fine's Nixon example
  - Lewis' problem of over-minimalization:

    $(A \vee B) > C$ entails $A > C$, $B > C$

5 Philosophical Assessment
5.1 Predictions

- **BUT:** the predictions made depend on the causal structure you assign to a concrete example

5 Philosophical Assessment
5.1 Predictions

King Ludwig of Bavaria likes to spend his weekends at Leoni Castle. Whenever the Royal Bavarian flag is up and the lights are on, the King is in the Castle. At the moment the lights are on, the flag is down, and the King is away.

(9) *If the flag were up, then the King would be in the castle.*

(10) *If the flag were up, then the light would be out.*

5 Philosophical Assessment
5.2 Is this about causation?

- conditionals exploit certain invariant relationships, certain dependencies
- what unifies these relationships is that the expressed dependency is one of manipulation and control:
  - *A stands in this relation to B if manipulating A will change B in a systematic way;*
  - *by manipulating A one can control B*
- an invariant relationship with these properties I call *causal* relation

5 Philosophical Assessment
5.2 Is this about causation?

It is a simple fact of basic math that if you add two natural numbers that are both even or uneven, the sum will be even, but if one of the numbers is even and the other uneven the sum is uneven. Suppose you're explaining this fact to some school kids and you have on the board 3 + 4 = 7. You say ...

(7) If the first number had been even, the result would have been even.
(8) If the result had been even, the first number would have been even.

5 Philosophical Assessment
5.5 What is causality?

"So beschouwd, is het causaliteitsbeginsel dus geen principe a priori; maar het is ook geen natuurwet en zeker geen conventie. We kunnen misschien het beste zeggen, dat het causaliteitsbeginsel op een mede door historische factoren bepaalde wijze uitdrukking geeft aan een algemeen-menselijke neiging, op grond van spontaan geäpperciëerde causale (en andere) verbanden een meer, en zo mogelijk alles, omvattende causale structuur op te bouwen, die ons in staat stelt het universum waarin wij leven als een kosmos te begrijpen" E. Beth

5 Philosophical Assessment
5.5 What is causality?

"From this perspective, the principle of causality is not a principle a priori; but it is also not a law of nature and certainly not a convention. Maybe, the best we can say is that the principle of causality expresses, partly determined by historical factors, a human tendency to build, based on spontaneously experienced causal (and other) relations, a more, and, if possible, all, including causal structure, which enables us to understand the universe we live in."