Explanatory Derivations in First-oder Logic

Francesca Poggiolesi
poggiolesi@gmail.com

CNRS, IHPST, UMR 8590

6 June 2023, Logic Colloquium, Milan, Italy

Aim of the talk

Explanation, the main object of this talk, has been one of the most intensely discussed topics of philosophy of science in the 20th century.

It is therefore useful to start with an orientation.

Several different types of explanations:

- explaining the meaning of a symbol,
- explaining a new concept to a child,
- explaining how to construct an Ikea furniture.

We only focus on deductive explanations why a phenomenon occurred/a proposition is true.

Causal Explanations

Causal Explanations → Causal Relation

Causal Explanations $\rightarrow$ Cause(s)[1]

$\vdots$

Phenomenon/Effect

[1] Both causes and effect are verified events or facts.

Causal Explanations → A cigarette lit in the forest

⋮

A fire in the forest

Causal Explanations → Burning of fossil fuels

⋮

Climate change

Although causal explanations are central in scientific inquiry and philosophy, logic has been argued to have a problematic relationship with causality.[1]

[1] E.g., see Scriven, M., The logic of cause, *Theory and Decision*, 2: 49-66, 1971. Note that quite recently there has been an opposite trend in the works of, e.g., Ibeling and Icard 2020, Moss, Ibeling, Icard, 2022.

As a result, as far as we know, there is no serious logical investigation of the concept of (causal) explanation.

Conceptual Explanations

Conceptual Explanations → Grounding Relation

Conceptual Explanations $\rightarrow$

Ground(s) or reason(s)[1]

$\vdots$

Conclusion

[1] Both grounds and conclusion are true items.

Conceptual Explanations → That animal being a female and that animal being a fox

⋮

That animal being a vixen

Conceptual Explanations $\rightarrow$

Jane being mathematically talented, Jane being a hard worker, Jane having logic interests....

$\vdots$

Jane being an ideal candidate for a post-doc in logic in Milan

Conceptual/Mathematical Explanations

For any two points x and y, there always exists a third z such that $I(xy)$ $=I(yz)=I(xz)$.

Conceptual/Mathematical Explanations $\rightarrow$ $\vdots$

For any two circles X and Y, one with center in x and radius xy, the other with center in y and radius xy, there exists a point z where they intersect and which is such that $I(xy)$ $=I(yz)=I(xz)$.

Bernard Bolzano, *Theory of Science*, 1837.

Conceptual explanations naturally invite logical analysis.

This is precisely the aim of this talk: to elaborate *a logical theory of conceptual explanations*.

On the one hand, this will allow us to introduce the notion of explanation in logic, which is so far being a great absentee of the logical literature.

On the other hand, this will enlighten our understanding of conceptual explanations.

As a methodology, we will rely on a dialogue between philosophy and logic.

Formal framework

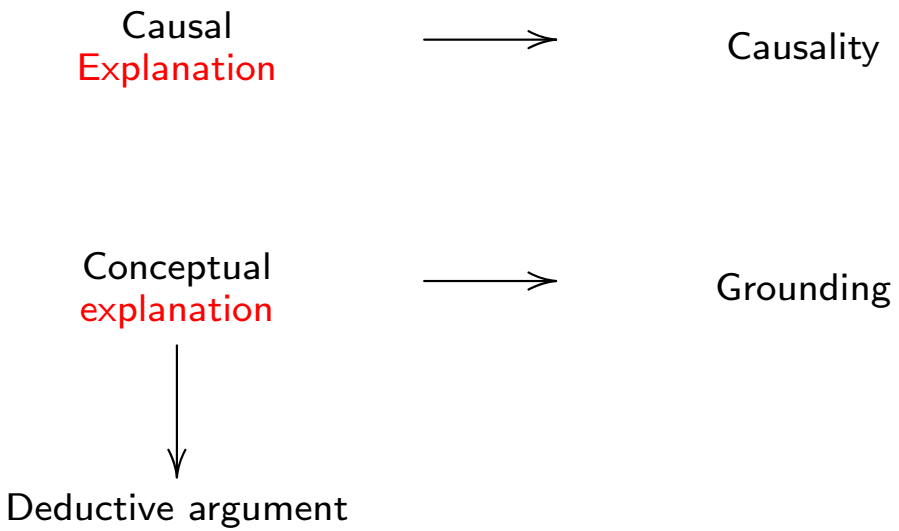
Causal Explanation $\longrightarrow$ Causality

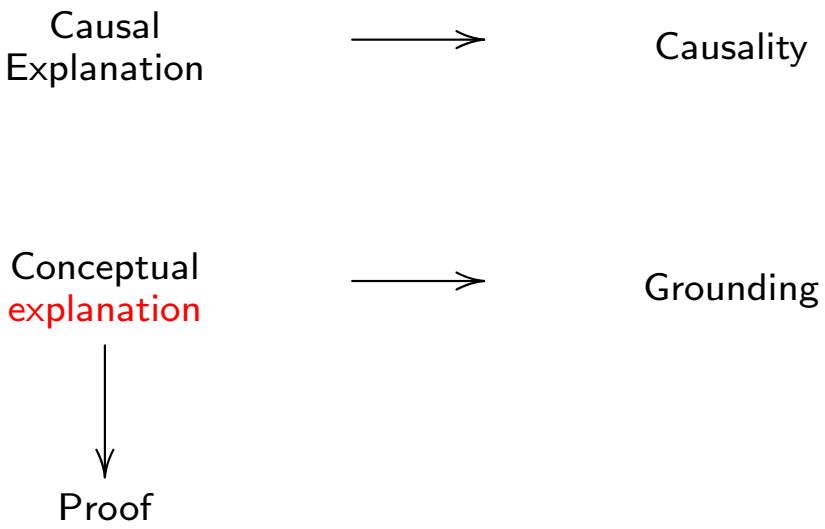
Causal Explanation $\longrightarrow$ Causality

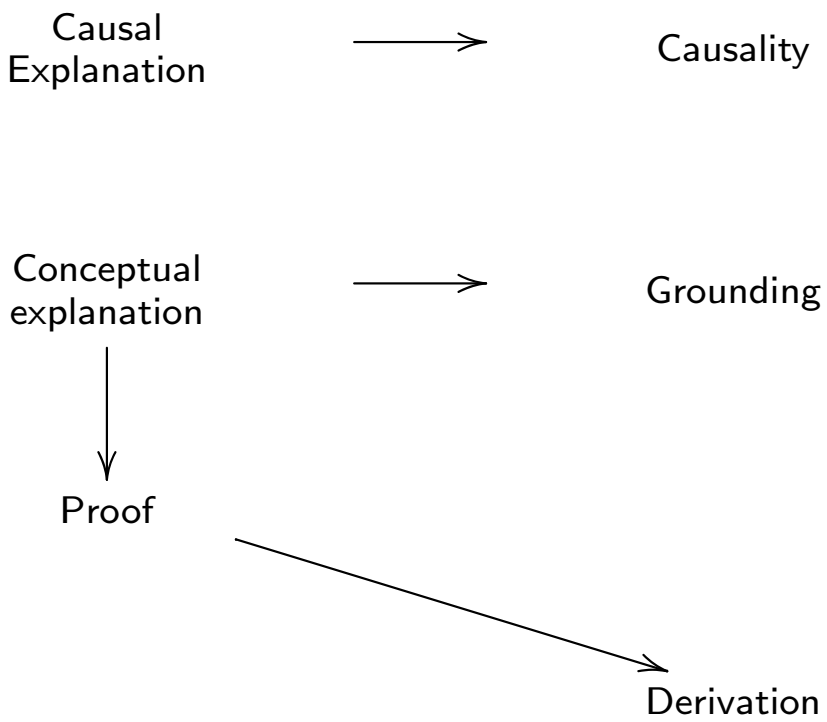
Conceptual explanation $\longrightarrow$ Grounding

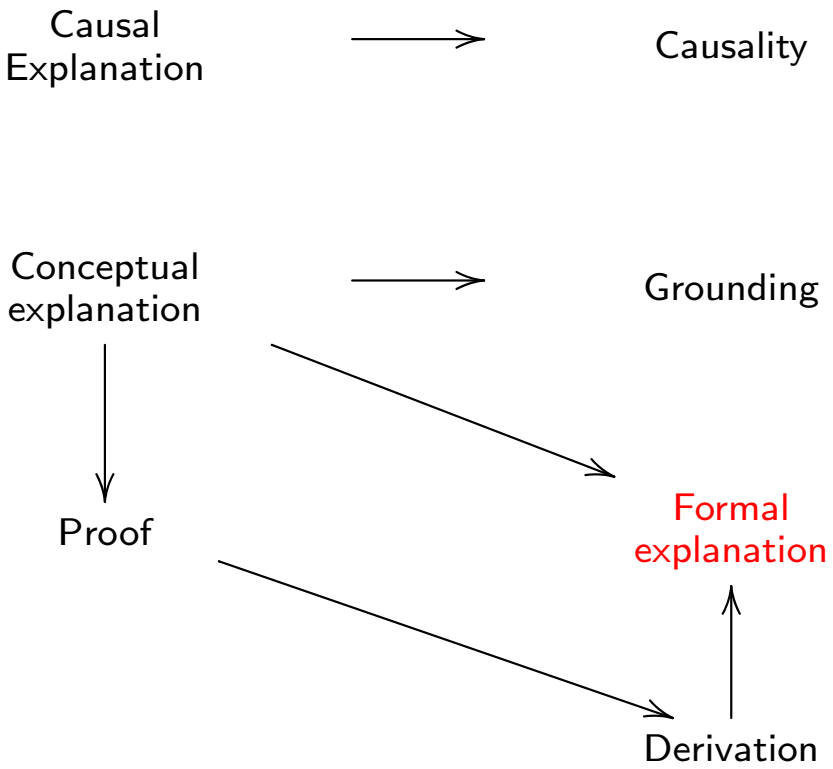
Causal Explanation $\longrightarrow$ Causality

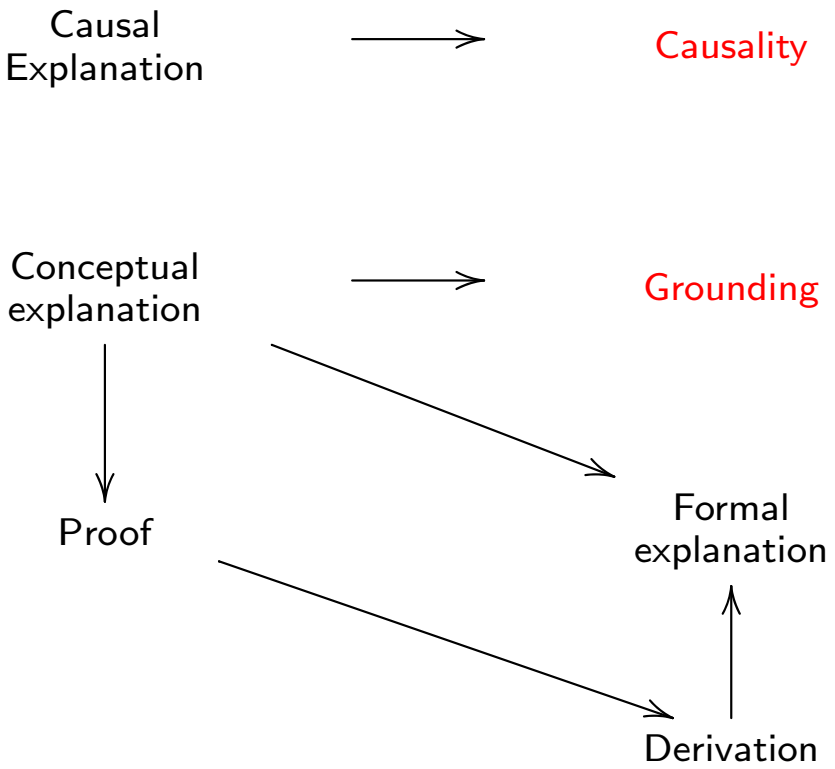
Conceptual explanation $\longrightarrow$ Grounding

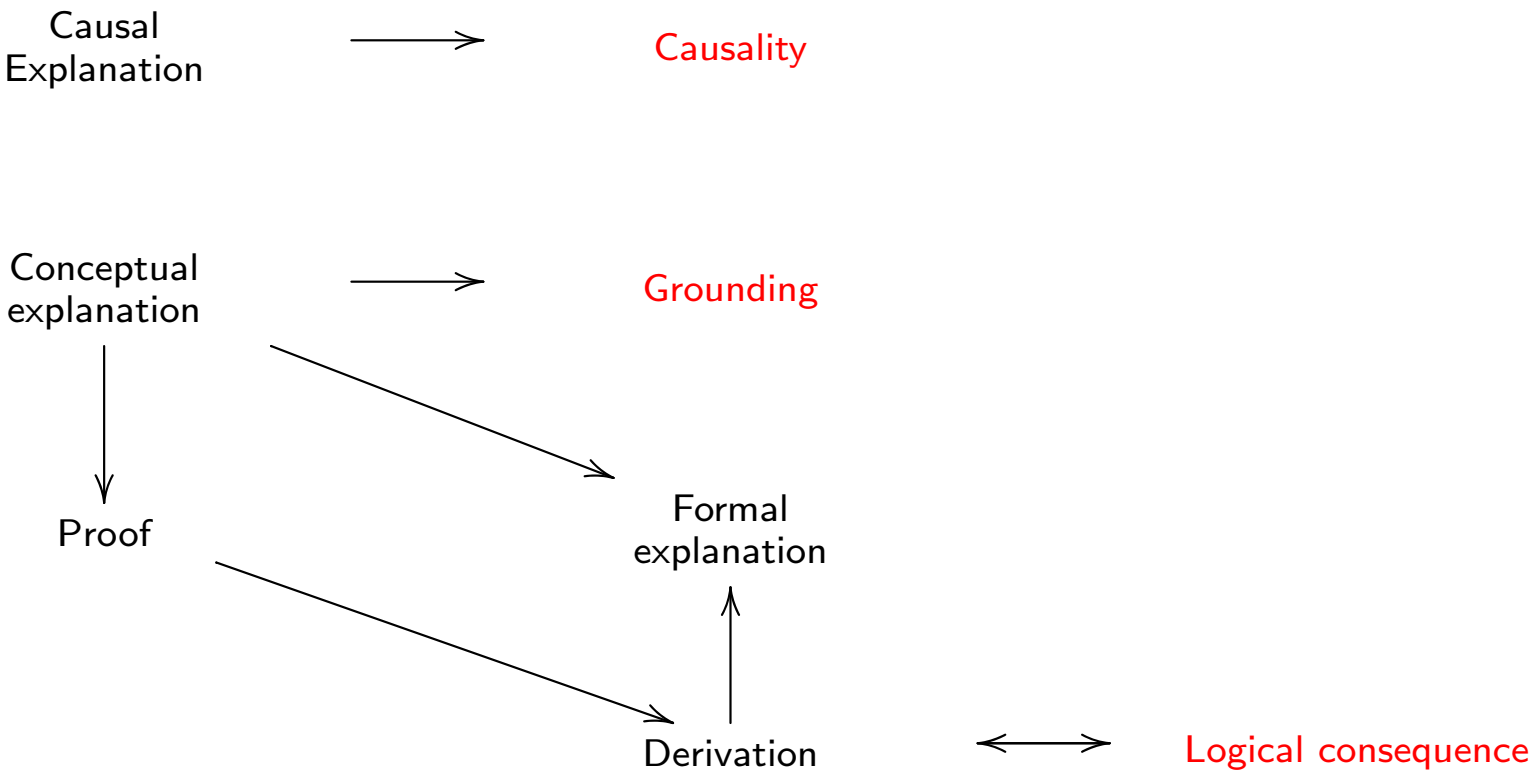

Causal
Explanation
Causality
Conceptual
explanation
Grounding
Deductive argument

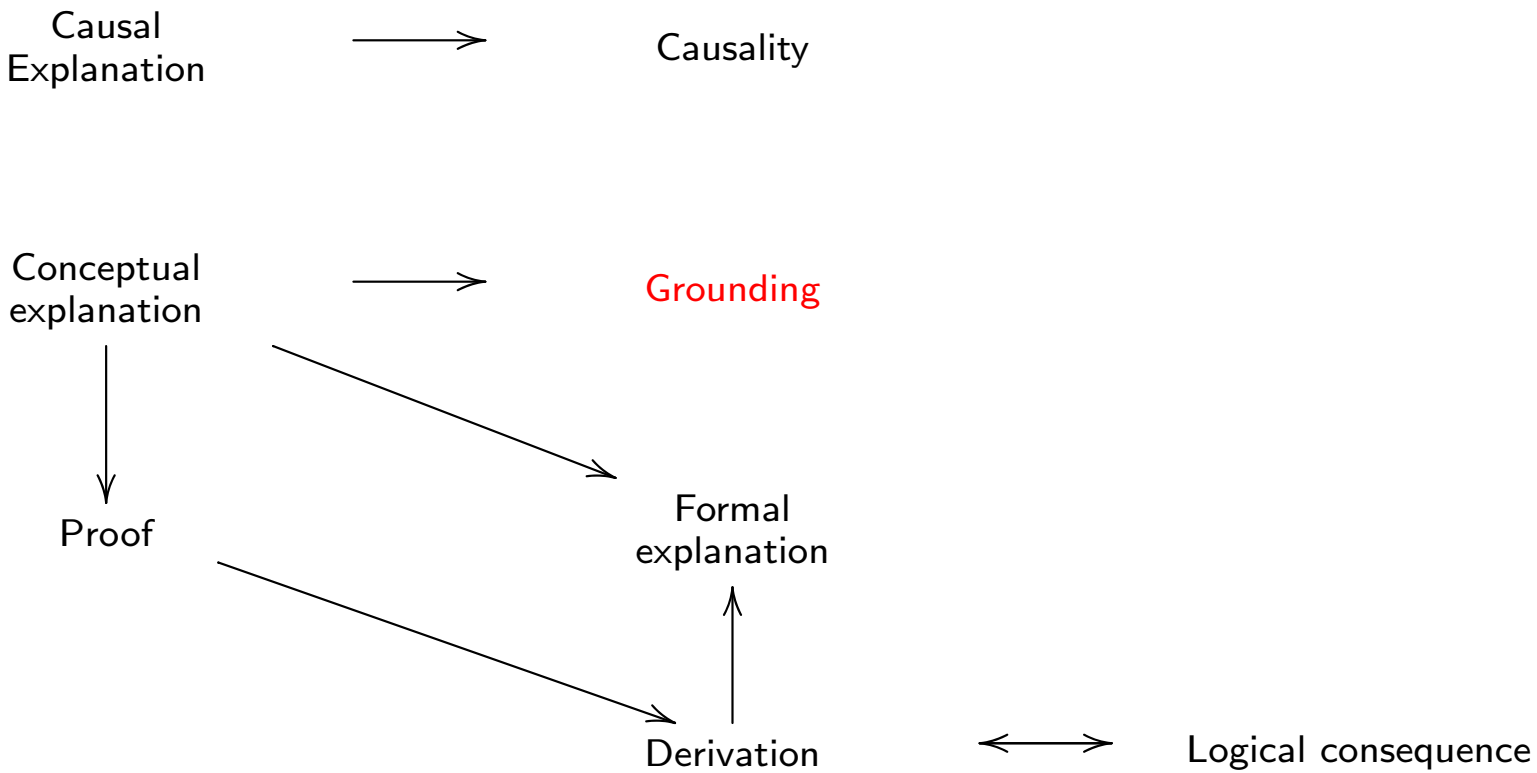

Causal
Explanation
Causality
Conceptual
explanation
Grounding
Proof

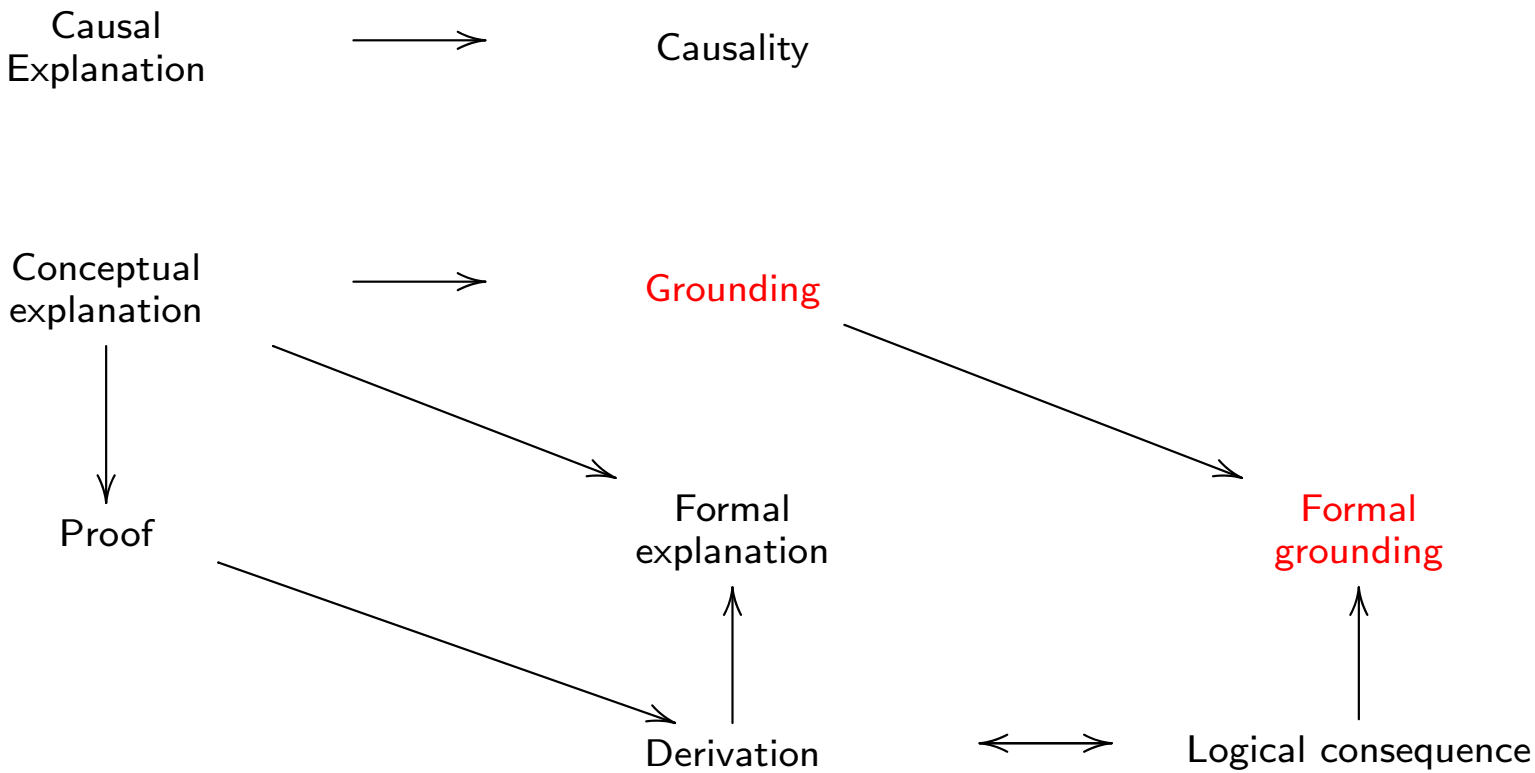

Causal
Explanation
Causality
Conceptual
explanation
Grounding
Proof
Derivation

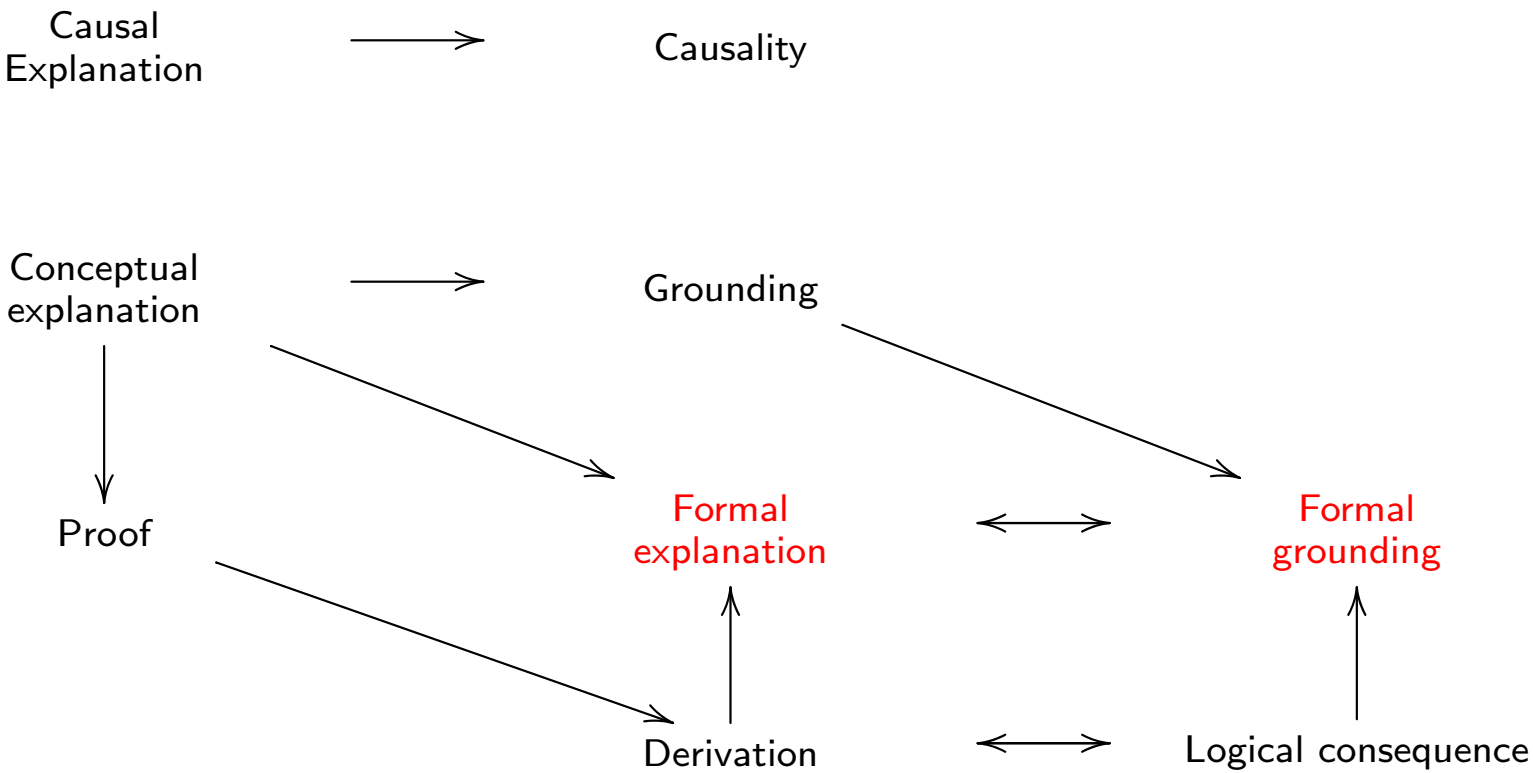

Causal
Explanation
Causality
Conceptual
explanation
Grounding
Proof
Formal
explanation
Derivation


Causal
Explanation
Causality
Conceptual
explanation
Grounding
Proof
Formal
explanation
Derivation


Causal Explanation
Causality
Conceptual explanation
Grounding
Proof
Formal explanation
Derivation
Logical consequence


Causal
Explanation
Causality
Conceptual
explanation
Grounding
Proof
Formal
explanation
Derivation
Logical consequence


Causal
Explanation
Causality
Conceptual
explanation
Grounding
Proof
Formal
explanation
Formal
grounding
Derivation
Logical consequence


Causal
Explanation
Causality
Conceptual
explanation
Grounding
Proof
Formal
explanation
Formal
grounding
Derivation
Logical consequence

Notation

Notation

Formal explanation

$\Vdash$

Notation

Formal explanation
$\Vdash$

Formal grounding
$\models$

Notation

For any multiset of formulas of *FOL* M, and any formula A

Formal explanation	Formal grounding
$M \Vdash A$	$M \Vdash A$

,

$M \Vdash A$ $M \Vdash A$

Complete/partial

$M \Vdash A$ $\qquad\qquad$ $M \Vdash A$

Complete formal explanation/complete grounding relation.

Reasons/conditions

$$M' \,|\, M \Vdash A \qquad\qquad M' \,|\, M \Vdash A$$

Formal explanation/formal grounding carry a distinction between reasons and conditions.

Billy is my brother and Suzy is my sister. I have a nephew. The reason why I have a nephew is that my sister has a child. Indeed a nephew is the son of my brother or my sister and my sister (Suzy) has a child. My brother could have had a child, but he does not. Hence my brother having a child is merely a potential reason of why I have a nephew.[1]

[1] There is a stringent parallel with causation, see Menzies and Beebee (2020).

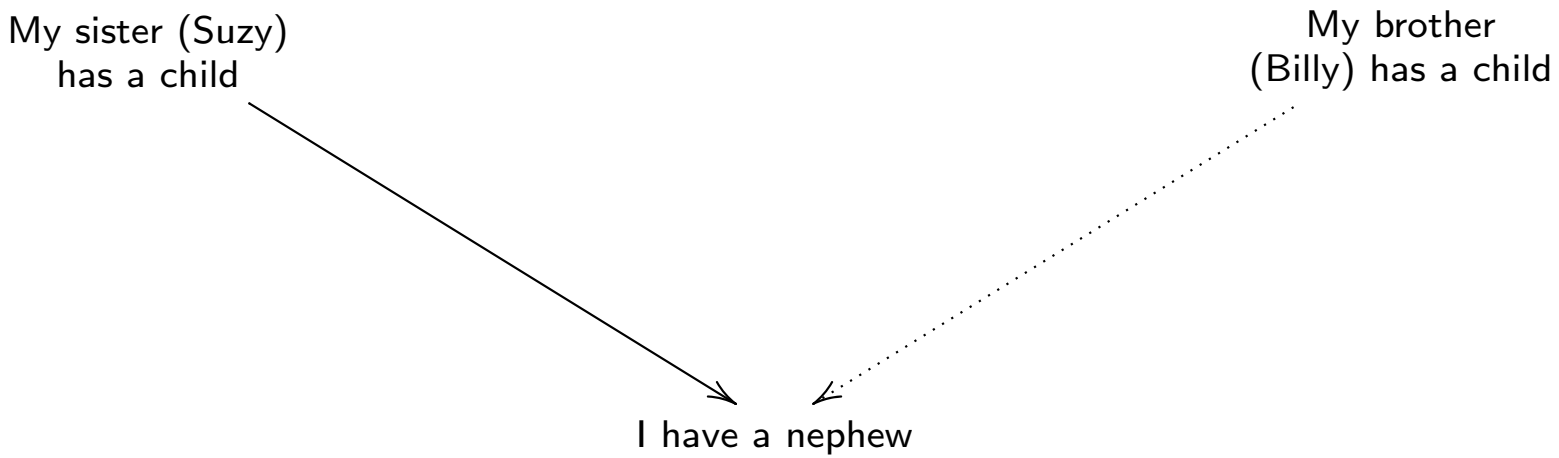
My sister (Suzy)
has a child

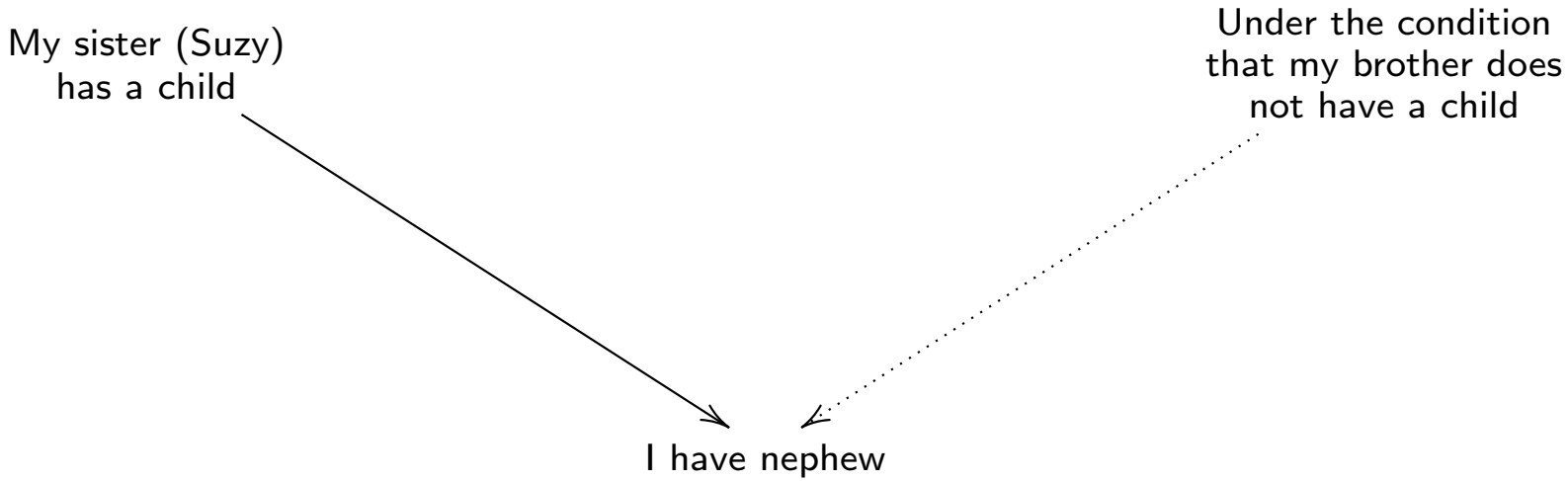
My brother
(Billy) has a child

My sister (Suzy)
has a child

My brother
(Billy) has a child

I have a nephew


My sister (Suzy)
has a child
My brother
(Billy) has a child
I have a nephew


My sister (Suzy)
has a child
Under the condition
that my brother does
not have a child
I have nephew

$M' \mid M \Vdash A$

$M' \mid M \Vdash A$

$M' \mid M \Vdash A$

$M' \mid M \Vdash A$

Logical analysis of formal grounding

$$M' \mid M \Vdash A$$

$$M' \mid M \Vdash A$$

What makes a consequence relation an explanatory one?[2]

[2] See F. Poggiolesi, On defining the notion of complete and immediate grounding, Synthese (2016). F. Poggiolesi and N. Francez, Towards a generalization of the logic of grounding, Theoria (2020). F. Genco, Formal Explanations as Logical Derivations (Francesco A. Genco). Journal of Applied Non-Classical Logics (2021).

$$M' \mid M \Vdash A$$

Which features define a grounding relation?

$$M' \mid M \Vdash A$$

Dependence.

$$M' \mid M \Vdash A$$

$$M' \mid F_1, ..., F_n \Vdash A$$

$$E_1, ..., E_m \mid F_1, ..., F_n \Vdash A$$

$$F_1, \ldots F_n \models A$$

Not only is the conclusion a consequence of the ground(s).

$$\neg F_1, \ldots F_n \models A$$

If the premisses were somehow changed

$$F_1, \ldots \neg F_n \models A$$

If the premisses were somehow changed

$$\neg F_1, \ldots \neg F_n \models A$$

If the premisses were somehow changed

$$\neg E_1, \ldots \neg E_m, \neg F_1, \ldots \neg F_n \models A$$

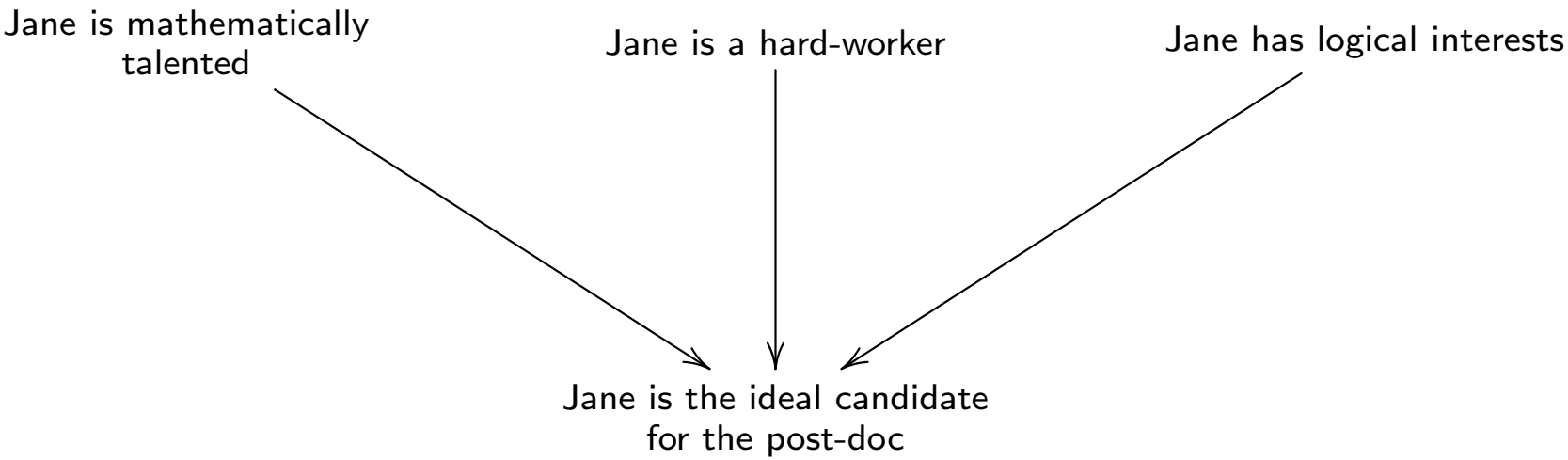
If the premisses were somehow changed plus the conditions,

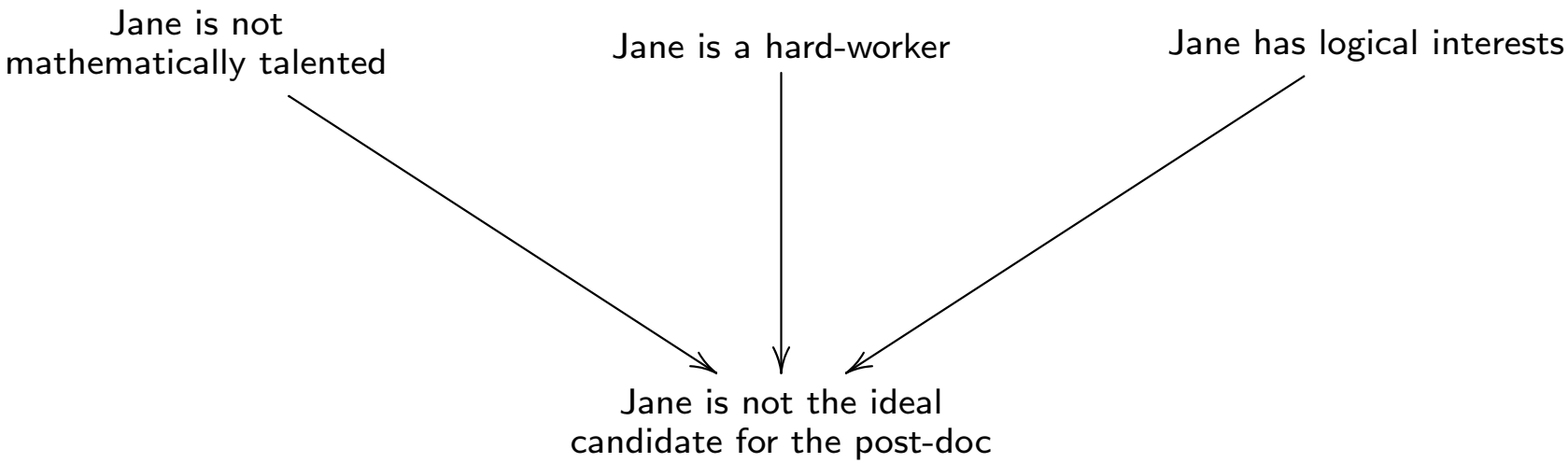
$$\neg E_1, \ldots \neg E_m, \neg F_1, \ldots \neg F_n \models \neg A$$

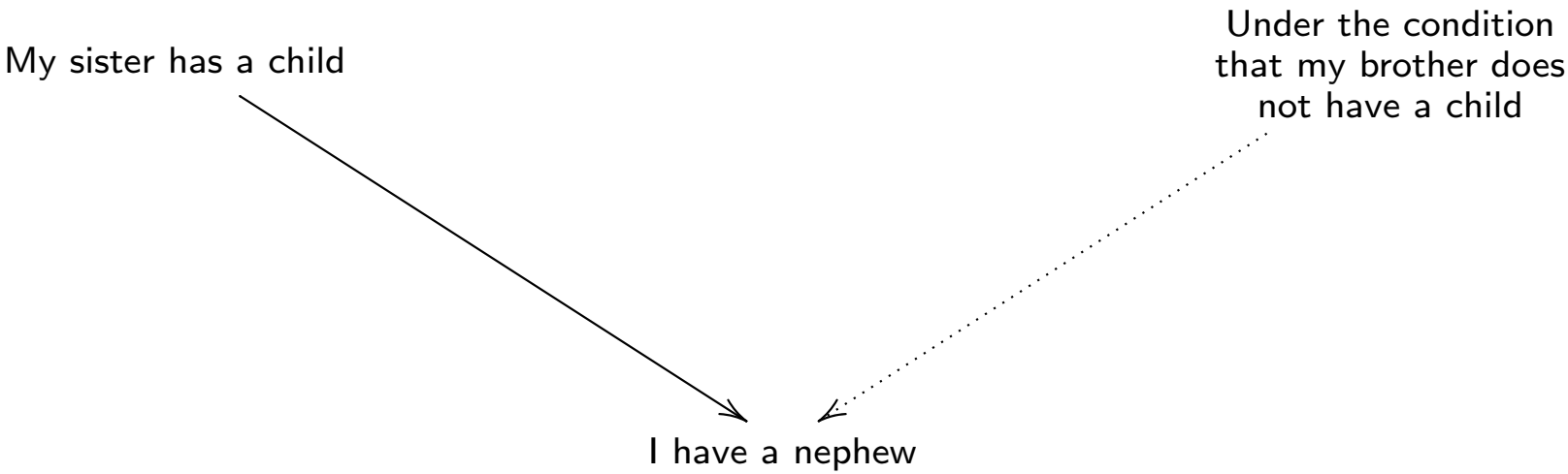
If the premisses were somehow changed plus the conditions, the change would affect the conclusion.

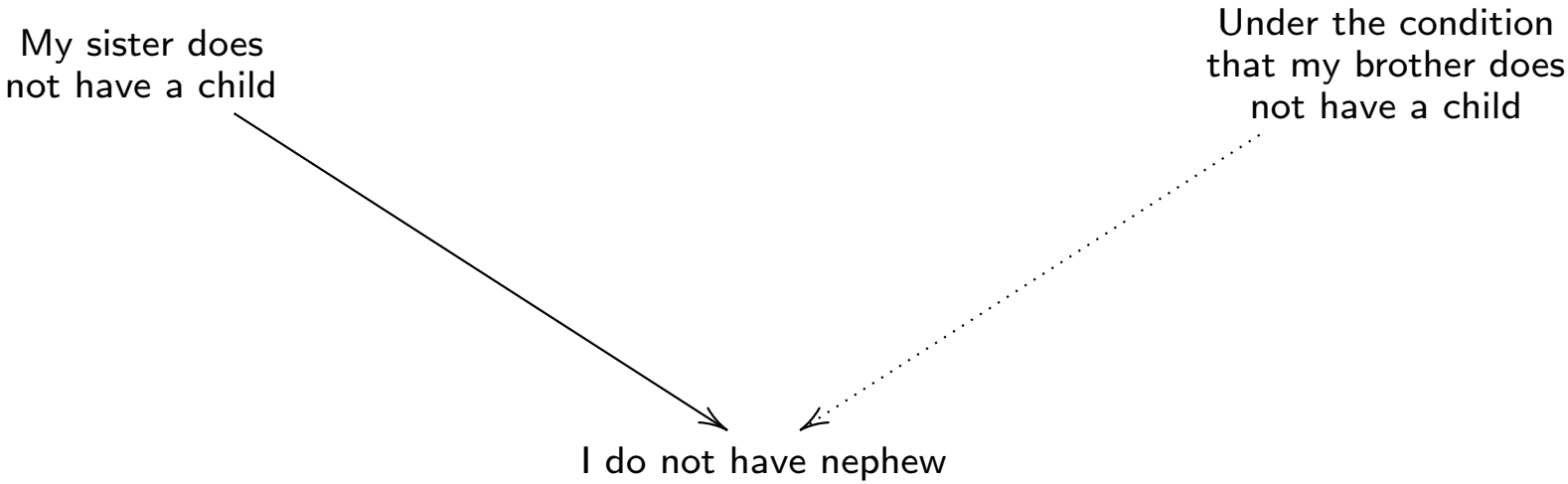
$$\neg E_1, \ldots \neg E_m, \neg F_1, \ldots \neg F_n \models \neg A$$

From the negation of some (even all) grounds + conditions, the negation of the conclusion follows.


Jane is mathematically talented
Jane is a hard-worker
Jane has logical interests
Jane is the ideal candidate for the post-doc


Jane is not
mathematically talented
Jane is a hard-worker
Jane has logical interests
Jane is not the ideal
candidate for the post-doc


My sister has a child
Under the condition
that my brother does
not have a child
I have a nephew


My sister does
not have a child
Under the condition
that my brother does
not have a child
I do not have nephew

$$M' \mid M \Vdash A$$

$$M' \mid M \Vdash A$$

A grounding relation is a dependence relation between premisses and the conclusion.

$$M' \mid M \Vdash A$$

Is that it?

$$M' \mid M \Vdash A$$

Explanatory relations are notoriously asymmetric relations.

$$F \Vdash A$$

In case of a unique ground the dependence boils down to an equivalence.

$$F \Leftrightarrow A$$

In case of a unique ground the dependence boils down to an equivalence.

For any two points x and y, there always exists a third z such that $l(xy) = l(yz) = l(xz)$.

$\Longleftrightarrow$

For any two circles X and Y, one with center in x and radius xy, the other with center in y and radius xy, there exists a point z where they intersect and which is such that $l(xy) = l(yz) = l(xz)$.

We need an ingredient which establishes what explains what.

$$M' \mid M \Vdash A$$

According to a long philosophical tradition, the ingredient is complexity!

$$M' \mid M \Vdash A$$

There is method, which is the most important, since we can use it to explain in any science. It consists in starting from the most general and simple things to move to the less general and more composed. Arnauld et Nicole, 1993, P. 228.

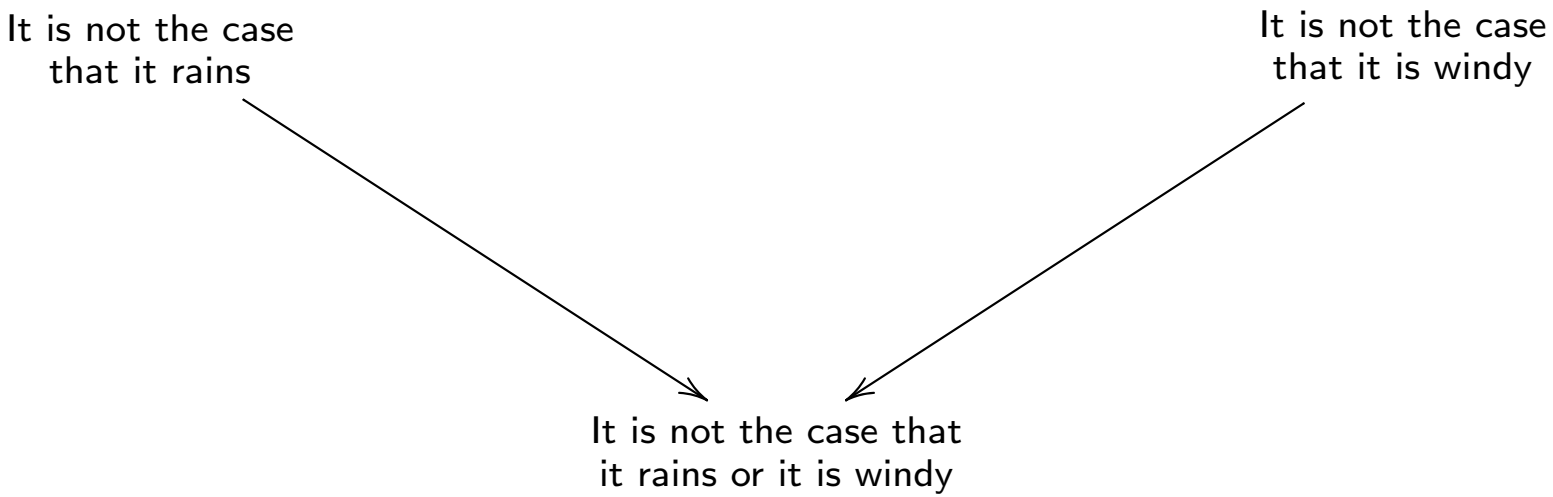
$$M' \mid M \Vdash A$$

Complexity standardly corresponds to logical complexity and related subformula.

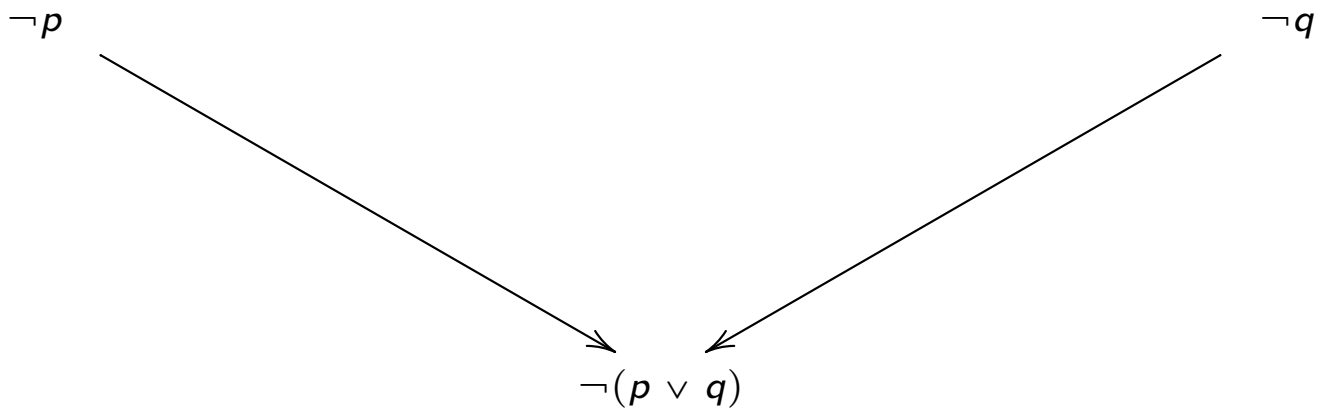
$$M' \mid M \Vdash A$$

Unfortunately, the formal grasp of the notion of explanation trough the notion of subformula turns out to be defective in several respects. (Adaptation of) Khale and Pulcini (2014).

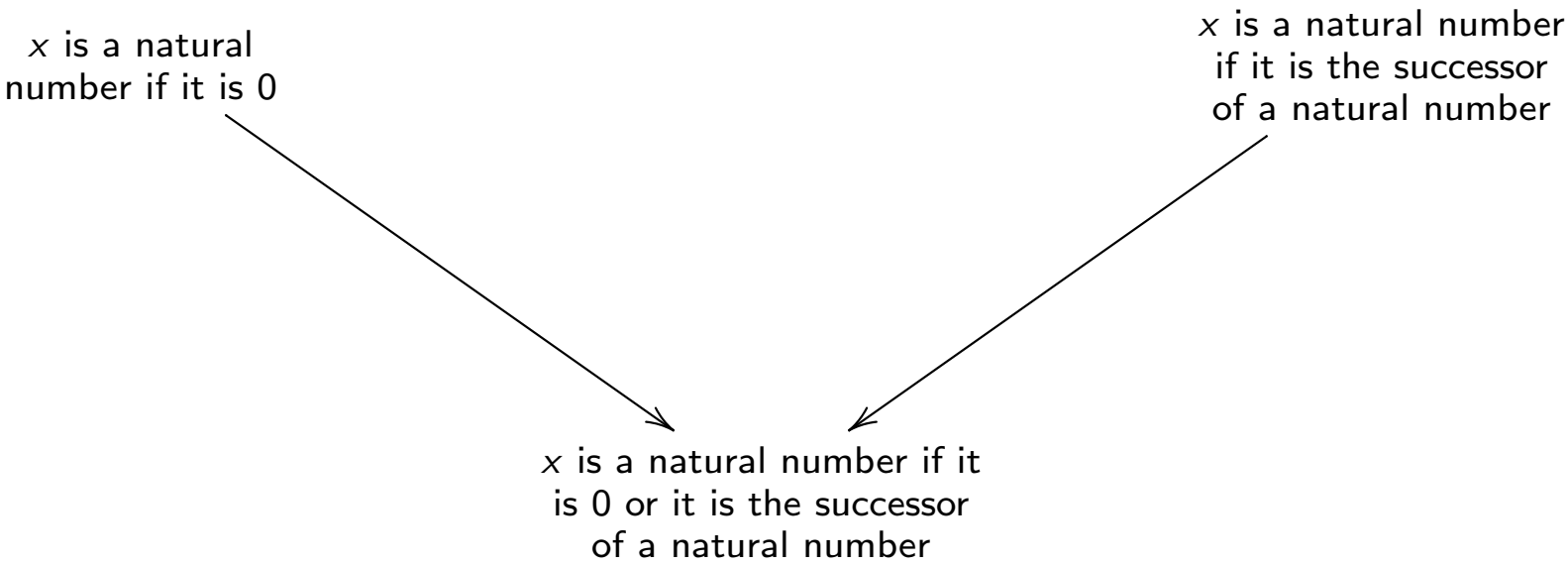
Counterexample 1



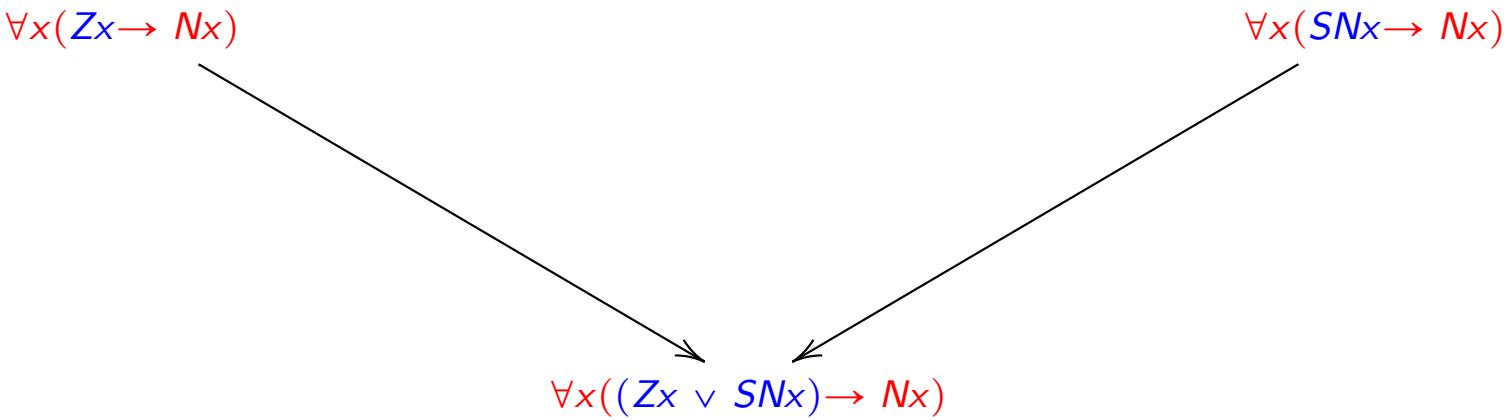
Counterexample 1



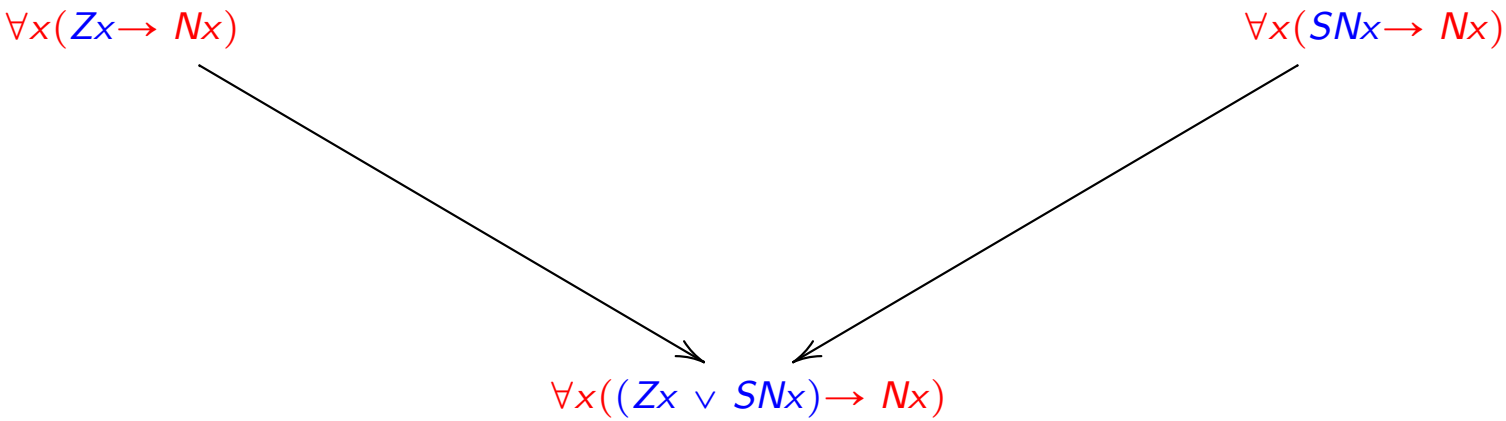
Counterexample 2



Counterexample 2



Counterexample 2



See *Deep inferences*, http://alessio.guglielmi.name/res/cos/.

We will enrich the notion of complexity/subformula so for them to fit in an explanatory framework.

We will consider the well-formed formulas of the language of first-order logic and divide them in a new hierarchy of complexity, g-complexity, which extends the standard one by taking into account 1.

In accordance with the new notion of g-complexity, we will define another relation of subformula, g-subformula, that extends the standard one by taking into account 2.

Level 0

Level 0 Pc, Qc, Rc, ...

Level 0 Pc, Qc, Rc, ..., $\neg Pc$, $\neg Qc$, $\neg Rc$, ...

Level 1

Level 0 Pc, Qc, Rc, ..., $\neg Pc$, $\neg Qc$, $\neg Rc$, ...

Level 1 $Pc \wedge Qc$,

Level 0 Pc, Qc, Rc, ..., $\neg Pc$, $\neg Qc$, $\neg Rc$, ...

Level 1 $Pc \vee Qc$,

Level 0 Pc, Qc, Rc, ..., $\neg Pc$, $\neg Qc$, $\neg Rc$, ...

Level 1 $Pc \circ Qc$,

Level 0 Pc, Qc, Rc, ..., $\neg Pc$, $\neg Qc$, $\neg Rc$, ...

Level 1 $Pc \circ Qc$, $\forall x Px$,

Level 0 Pc, Qc, Rc, ..., $\neg Pc$, $\neg Qc$, $\neg Rc$, ...

Level 1 $Pc \circ Qc$, $\exists x Px$,

Level 0 Pc, Qc, Rc, ..., $\neg Pc$, $\neg Qc$, $\neg Rc$, ...

Level 1 $Pc \circ Qc$, $\odot xPx$,

Level 0 Pc, Qc, Rc, ..., $\neg Pc$, $\neg Qc$, $\neg Rc$, ...

Level 1 $Pc \circ Qc,\ \neg(Pc \circ Qc),\ \odot xPx,$

Level 0 $Pc,\ Qc,\ Rc,\ ...,\ \neg Pc,\ \neg Qc,\ \neg Rc,\ ....$

Level 1 $Pc \circ Qc,\ \neg(Pc \circ Qc),\ \odot xPx,\ \neg \odot xPx,$

Level 0 $Pc,\ Qc,\ Rc,\ ...,\ \neg Pc,\ \neg Qc,\ \neg Rc,\ ....$

$\neg\neg\neg Pc$, ...

Level 1 $\neg\neg Pc$, ..., $Pc \circ Qc$, $\neg(Pc \circ Qc)$, $\odot xPx$, $\neg \odot xPx$, ...

Level 0 Pc, Qc, Rc, ..., $\neg Pc$, $\neg Qc$, $\neg Rc$, ...

$\vdots$

Level 2

$\neg\neg\neg Pc$, ...

Level 1 $\neg\neg Pc$, ..., $Pc \circ Qc$, $\neg(Pc \circ Qc)$, $\odot xPx$, $\neg \odot xPx$, ...

Level 0 Pc, Qc, Rc, ..., $\neg Pc$, $\neg Qc$, $\neg Rc$, ...

$\vdots$

$\neg\neg\neg\neg\neg Pc$

Level 2 $\quad \neg\neg\neg\neg Pc,\ \neg\neg(Pc \circ Qc),\ \forall y \forall x(P(x,y)),\ Rc \vee (Pc \wedge Qc)$

$\neg\neg\neg Pc\ ,\ \ ...$

Level 1 $\quad \neg\neg Pc,\ \ ...,\ Pc \circ Qc,\ \neg(Pc \circ Qc),\ \odot xPx,\ \neg \odot xPx,\ \ldots$

Level 0 $\quad Pc,\ Qc,\ Rc,\ ...,\ \neg Pc,\ \neg Qc,\ \neg Rc,\ ....$

$\vdots$

$\neg\neg\neg\neg\neg Pc$

Level 2 $\quad \neg\neg\neg\neg Pc,\ \neg\neg(Pc \circ Qc),\ \forall y \forall x(P(x,y)),\ Rc \vee (Pc \wedge Qc)$

$\neg\neg\neg Pc\ ,\ \ ...$

Level 1 $\quad \neg\neg Pc,\ \ ...,\ Pc \circ Qc,\ \neg(Pc \circ Qc),\ \odot xPx,\ \neg \odot xPx,\ \ldots$

Level 0 $\quad Pc,\ Qc,\ Rc,\ ...,\ \neg Pc,\ \neg Qc,\ \neg Rc,\ ....$

$\vdots$

$\neg\neg\neg\neg\neg Pc$

Level 2 $\quad \neg\neg\neg\neg Pc,\ \neg\neg(Pc \circ Qc),\ \forall y \forall x(P(x,y)),\ Rc \vee (Pc \wedge Qc)$

$\neg\neg\neg Pc\ ,\ \ ...$

Level 1 $\quad \neg\neg Pc,\ \ ...,\ Pc \circ Qc,\ \neg(Pc \circ Qc),\ \odot xPx,\ \neg \odot xPx,\ \ldots$

Level 0 $\quad Pc,\ Qc,\ Rc,\ ...,\ \neg Pc,\ \neg Qc,\ \neg Rc,\ ....$

$\vdots$

$\neg\neg\neg\neg\neg Pc$

Level 2 $\quad \neg\neg\neg\neg Pc,\ \neg\neg(Pc \circ Qc),\ \forall y \forall x(P(x,y)),\ \ Rc \vee (Qc \wedge Pc)$

$\neg\neg\neg Pc\ ,\ \ ...$

Level 1 $\quad \neg\neg Pc,\ \ ...,\ Pc \circ Qc,\ \neg(Pc \circ Qc),\ \odot xPx,\ \neg \odot xPx,\ \ldots$

Level 0 $\quad Pc,\ Qc,\ Rc,\ ...,\ \neg Pc,\ \neg Qc,\ \neg Rc,\ ....$

$\vdots$

$\neg\neg\neg\neg\neg Pc$

Level 2 $\quad \neg\neg\neg\neg Pc,\ \neg\neg(Pc \circ Qc),\ \forall y \forall x(P(x,y)),\ \ Rc \vee (Qc \wedge Pc)$

$\neg\neg\neg Pc\ ,\ \ ...$

Level 1 $\quad \neg\neg Pc,\ \ ...,\ Pc \circ Qc,\ \neg(Pc \circ Qc),\ \odot xPx,\ \neg \odot xPx,\ \ldots$

Level 0 $\quad Pc,\ Qc,\ Rc,\ ...,\ \neg Pc,\ \neg Qc,\ \neg Rc,\ ....$

$\vdots$

$\neg\neg\neg\neg\neg Pc$

Level 2 $\quad \neg\neg\neg\neg Pc,\ \neg\neg(Pc \circ Qc),\ \forall x \forall y(P(x,y)),\ \ Rc \vee (Qc \wedge Pc)$

$\neg\neg\neg Pc\ ,\ \ ...$

Level 1 $\quad \neg\neg Pc,\ \ ...,\ Pc \circ Qc,\ \neg(Pc \circ Qc),\ \odot xPx,\ \neg \odot xPx,\ \ldots$

Level 0 $\quad Pc,\ Qc,\ Rc,\ ...,\ \neg Pc,\ \neg Qc,\ \neg Rc,\ ....$

$\vdots$

$\neg\neg\neg\neg\neg Pc$

Level 2 $\quad \neg\neg\neg\neg Pc,\ \neg\neg(Pc \circ Qc),\ \forall k \forall w(P(k,w)),\ \ Rc \vee (Qc \wedge Pc)$

$\neg\neg\neg Pc\ ,\ \ ...$

Level 1 $\quad \neg\neg Pc,\ \ ...,\ Pc \circ Qc,\ \neg(Pc \circ Qc),\ \odot xPx,\ \neg \odot xPx,\ \ldots$

Level 0 $\quad Pc,\ Qc,\ Rc,\ ...,\ \neg Pc,\ \neg Qc,\ \neg Rc,\ ....$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

What are its (immediate) subformulas?

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

It depends on which part of the formula we focus on.

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow Qx \wedge Ry)$$

It depends on which part of the formula we focus on.

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

$\exists(Sx \wedge Tx)$

$\forall x \forall y(Px \rightarrow Qx \wedge Ry)$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

$\exists(Sx \wedge Tx)$

$\forall x \forall y(Px \rightarrow Qx \wedge Ry)$

$\neg\exists(Sx \wedge Tx)$

$\neg\forall x \forall y(Px \rightarrow Qx \wedge Ry)$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

Now suppose we focus on another part of the formula.

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

Now suppose we focus on another part of the formula.

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

A[]

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

$$\mathsf{A}[Qx \wedge Ry]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

$$\mathrm{A}[Qx \wedge Ry]$$

$$\mathrm{A}[Qx]$$

$$\mathrm{A}[Ry]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow Qx \wedge Ry)$$

$$\mathrm{A}[Qx \wedge Ry]$$

$$\mathrm{A}[Qx]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow Qx)$$

$$\mathrm{A}[Ry]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow Ry)$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

$$A[Qx \wedge Ry]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

$$A[Qx \wedge Ry]$$

$$\mathrm{A}[\neg Qx]$$

$$\mathrm{A}[\neg Ry]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow Qx \wedge Ry)$$

$$A[Qx \wedge Ry]$$

$$\mathrm{A}[\neg Qx]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow \neg Qx)$$

$$\mathrm{A}[\neg Ry]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow \neg Ry)$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

$$A[Qx \wedge Ry]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow Qx \wedge Ry)$$

$$A[Qx \wedge Ry]$$

$$\neg A[Qx]$$

$$\neg A[Ry]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow Qx \wedge Ry)$$

$$A[Qx \wedge Ry]$$

$$\neg A[Qx]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow \neg Qx)$$

$$\neg A[Ry]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow \neg Ry)$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y(Px \rightarrow Qx \wedge Ry)$$

$$A[Qx \wedge Ry]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow Qx \wedge Ry)$$

$$A[Qx \wedge Ry]$$

$$\neg A[\neg Qx] \qquad \neg A[\neg Ry]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow Qx \wedge Ry)$$

$$A[Qx \wedge Ry]$$

$$\neg A[\neg Qx]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow \neg Qx)$$

$$\neg A[\neg Ry]$$

$$\exists(Sx \wedge Tx) \vee \forall x \forall y (Px \rightarrow \neg Ry)$$

$$M' \mid M \Vdash A$$

$$M' \mid M \Vdash A$$

A grounding relation is a dependence relation where the grounds and conditions are g-subformulas of the conclusion.

For a more precise definition...

Definition

If $A[-]$ is a formula with a context, the *scope of a context*, $SC(D)$, is defined inductively in the following way:

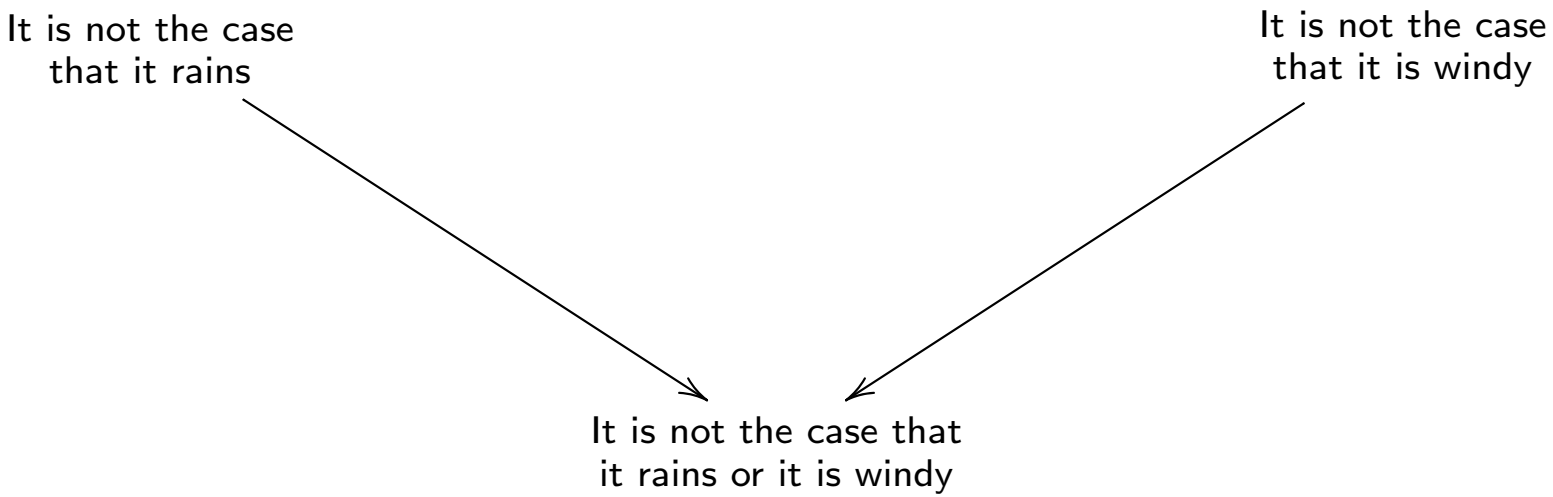
- if $A[]= []$, then $SC(A) = \varnothing$,
- if $A[]= E \circ F[]$, then $SC(A) = SC(F)$,
- if $A[]= \forall x F[]$, then $SC(A) = \forall x.^{+} SC(E)$,
- if $A[]= \exists x F[]$, then $SC(A) = \exists x.^{+} SC(E)$,
- if $A[]= \neg F[]$, then $SC(A) = (SC(F))^{*}$, where $*$ stands for swap the polarities.

Definition

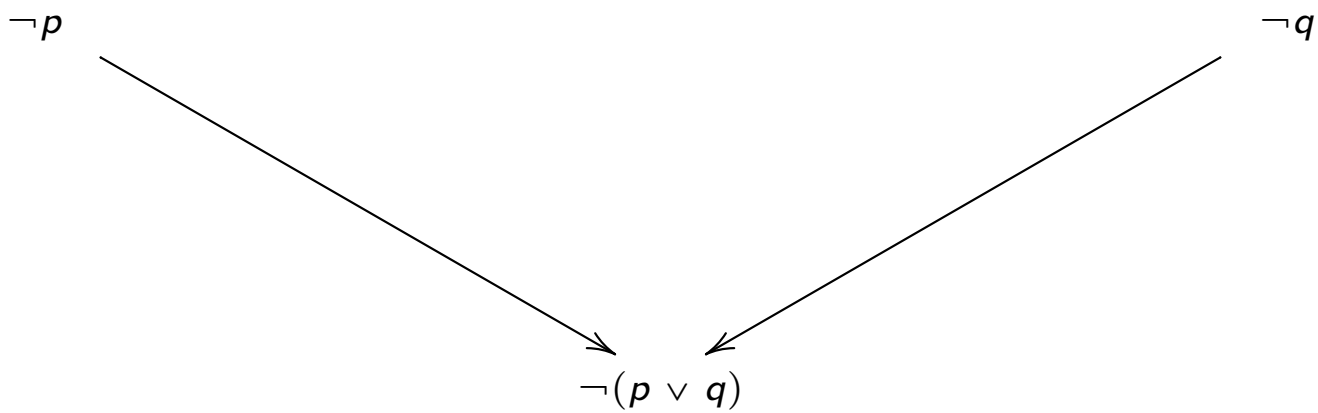
For any multiset of formulas (which could be empty) $M' = \{A[B_1], ..., A[B_n]\}$ and for any multiset of formulas $M = \{A[D_1], ..., A[D_m]\}$, under the condition that $M^{\perp}$, M completely and immediate ground $A[C]$, $M' \mid M \mid\!\doteq A[C]$,
if an only if, for any $E[-]$ such that $SC(E) = SC(A)$, we have:

- $M/E \vdash E[C]$,
- for some non empty (possibly non proper) submultiset M'' of M, we have that $M'^{\perp}/E, M''^{\perp}/E, M^{-}/E \vdash E[C]^{\perp}$.
- $M \cup M'$ are immediate (and complete) subformulas of $A[C]$.

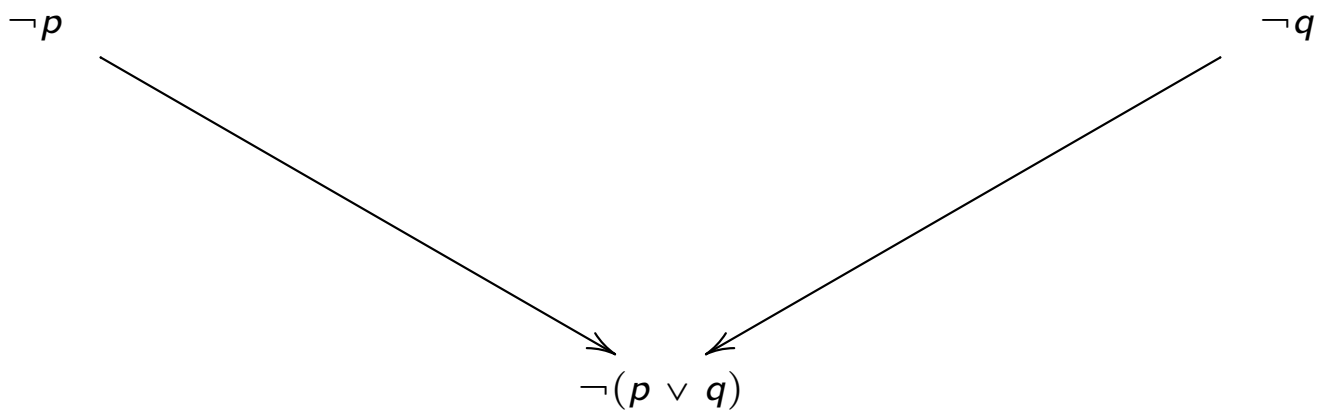
Example 1


It is not the case
that it rains
It is not the case
that it is windy
It is not the case that
it rains or it is windy

Example 1

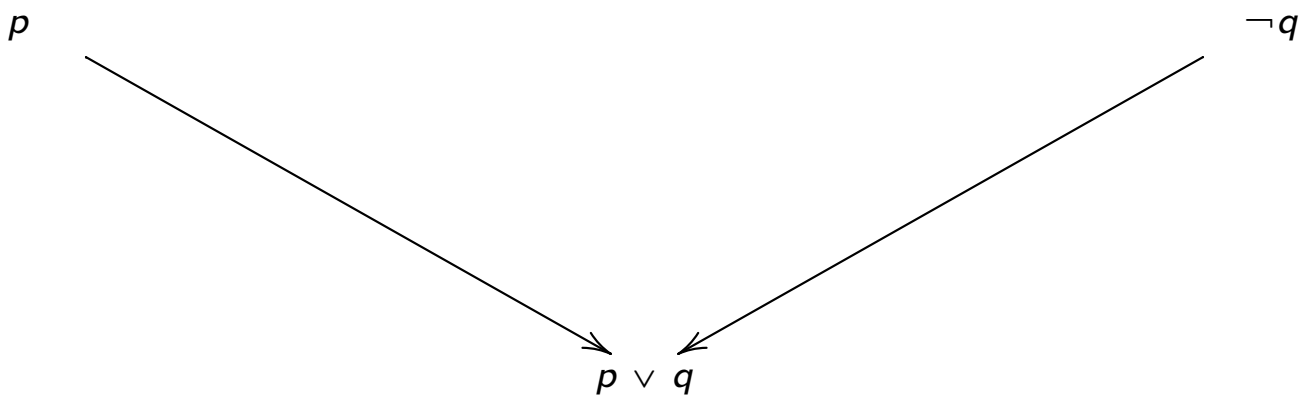


Example 1



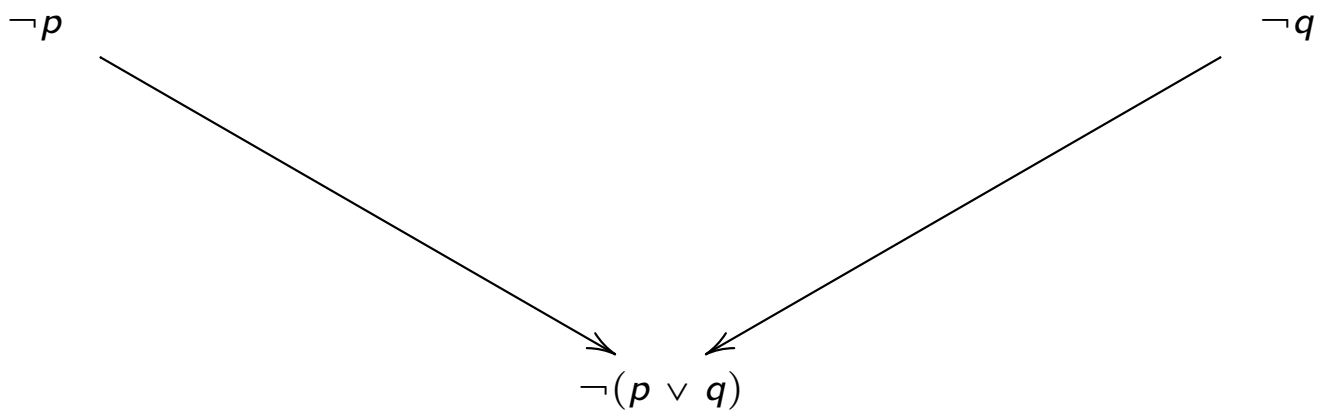
$\neg p, \neg q \models \neg(p \vee q)$

Example 1



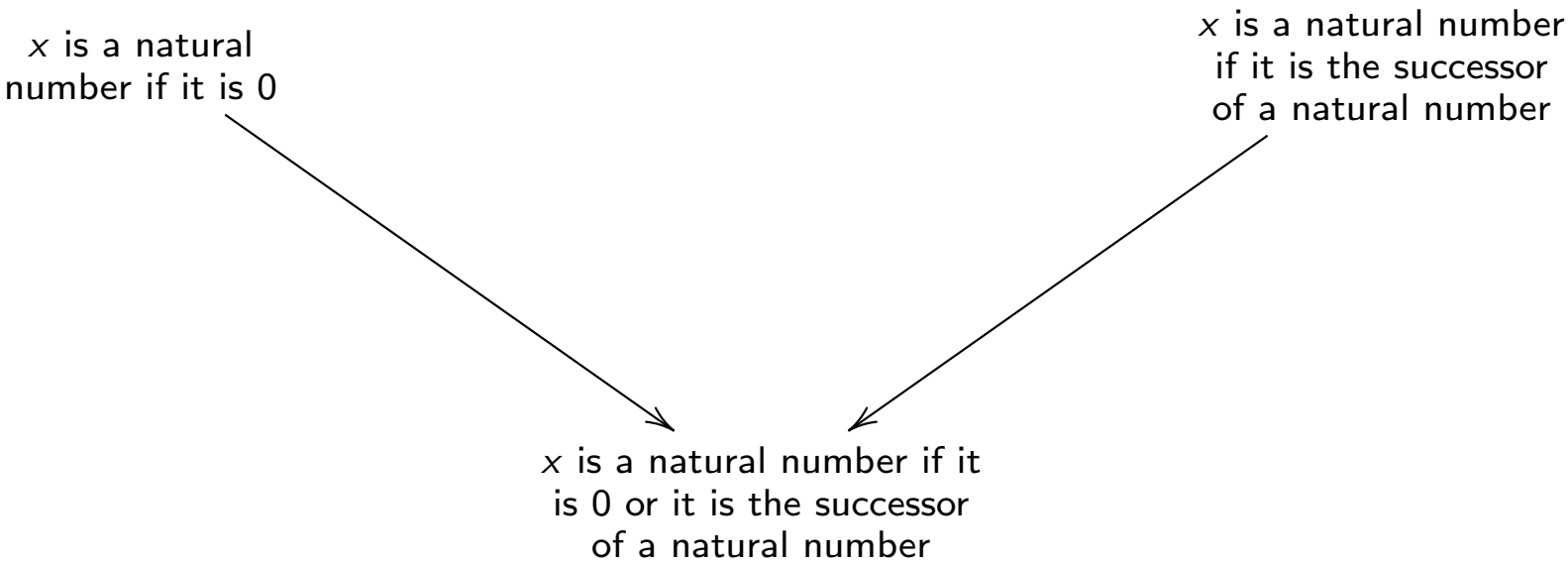
$p, \neg q \models p \vee q$

Example 1

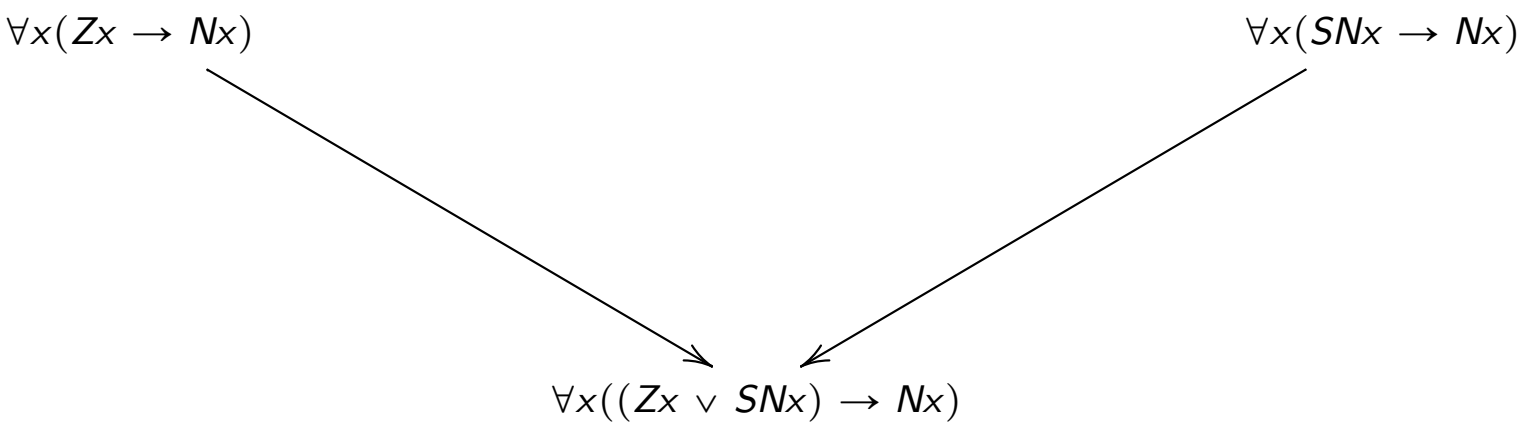


$\neg p$ and $\neg q$ are g-subformulas of $\neg(p \vee q)$

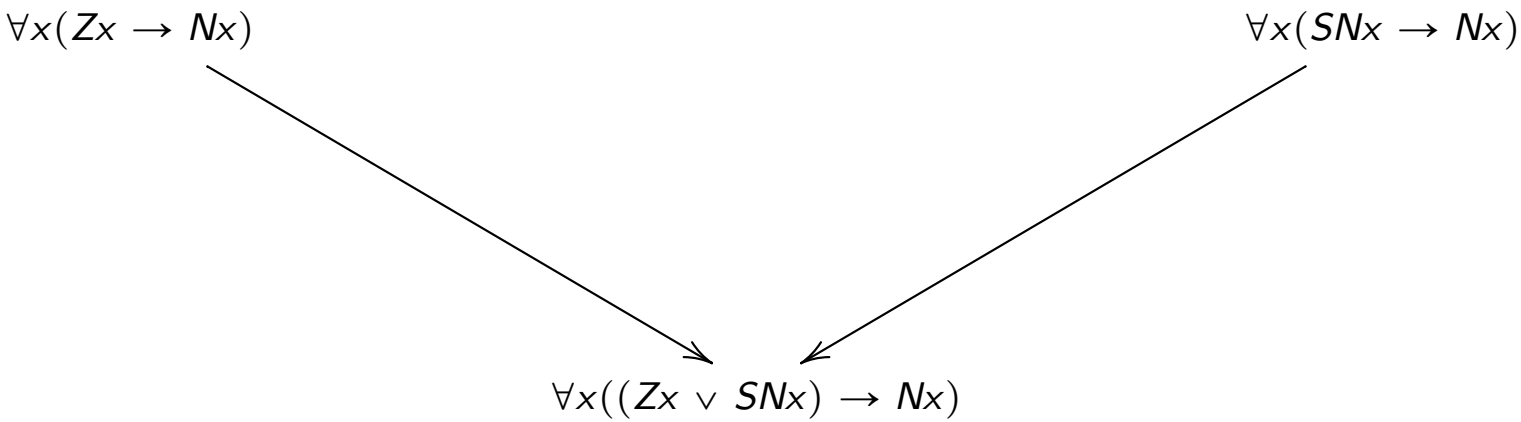
Example 2


x is a natural number if it is 0
x is a natural number if it is the successor of a natural number
x is a natural number if it is 0 or it is the successor of a natural number

Example 2

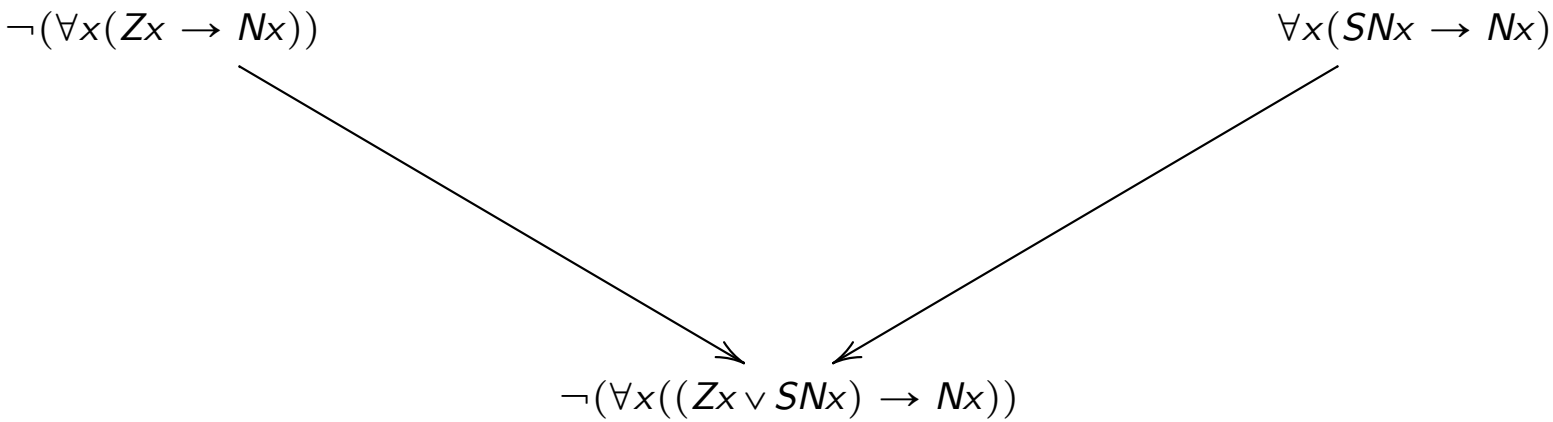


Example 2



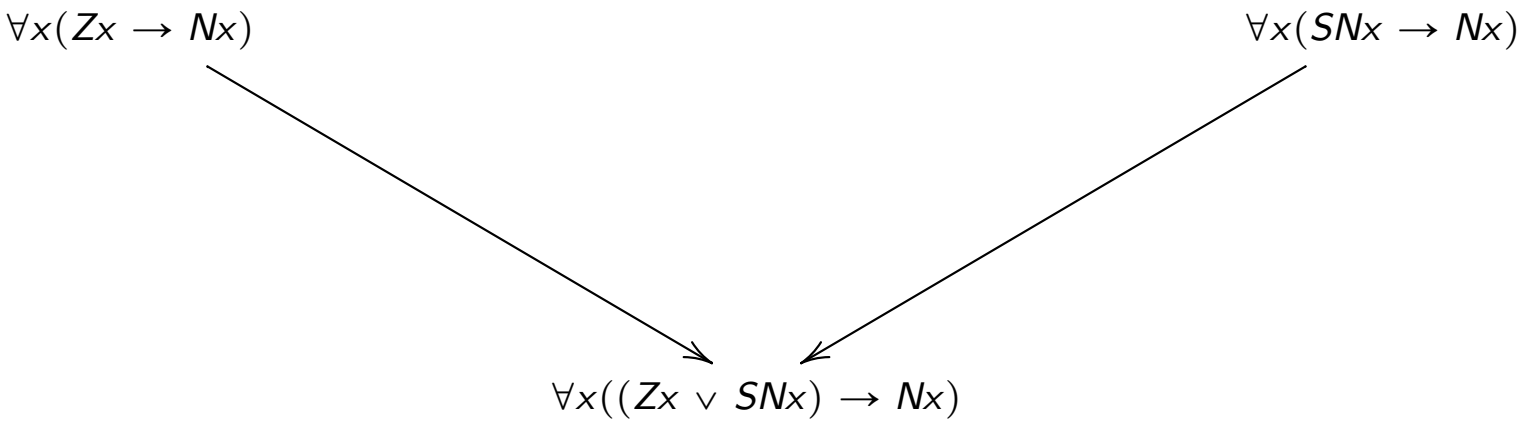
$\forall x(Zx \to Nx), \forall x(SNx \to Nx) \models \forall x(Zx \vee SNx \to Nx)$

Example 2



$\neg(\forall x(Zx \to Nx)), \forall x(SNx \to Nx) \models \neg(\forall x(Zx \vee SNx \to Nx))$

Example 2



$\forall x(Zx \to Nx)$ and $\forall x(SNx \to Nx)$ are g-subformulas of $\forall x(Zx \vee SNx \to Nx)$

Logical analysis of formal explanation

$$M' \mid M \Vdash A$$

$$M' \mid M \Vdash A$$

$$M' \mid M \Vdash A$$

We need to introduce (explanatory) rules that define a formal explanation.

$$M' \mid M \Vdash A$$

The rules will be such that their premisses are the complete grounds of the conclusion.

$$M' \mid M \Vdash A$$

Hence they will reflect at the proof-theoretical level the features of the grounding relation.

We will work with the sequent calculus extending previous results obtained in natural deduction.[2]

[2] See Poggiolesi, F., *On constructing a logic for the notion of complete and immediate formal grounding*, Synthese, 2018. Genco, F. *What Stands Between Grounding Rules and Logical Rules Is the Excluded Middle*, RSL, forthcoming.

Gfcl

Initial Sequents $p, M \Rightarrow M, p$

Propositional Logical Rules

$$\frac{M \Rightarrow N, A}{\neg A, M \Rightarrow N}\ \neg L \qquad \frac{A, M \Rightarrow N}{M \Rightarrow N, \neg A}\ \neg R$$

$$\frac{A_0, A_1, M \Rightarrow N}{A_0 \wedge A_1, M \Rightarrow N}\ \wedge L \qquad \frac{M \Rightarrow N, A \qquad M \Rightarrow N, B}{M \Rightarrow N, A \wedge B}\ \wedge R$$

First-order Logical Rules

$$\frac{\forall x A, A(x/t), M \Rightarrow N}{\forall x A, M \Rightarrow N}\ \forall L \qquad \frac{M \Rightarrow N, A(y)}{M \Rightarrow N, \forall x A}\ \forall R$$

where the y does not occur neither in M nor in N.

Gfcl

Initial Sequents $p, M \Rightarrow M, p$ $\qquad$ $\mathrm{M} \Rightarrow N \mid$

Propositional Logical Rules

$$\frac{M \Rightarrow N, A}{\neg A, M \Rightarrow N}\ \neg L \qquad \frac{A, M \Rightarrow N}{M \Rightarrow N, \neg A}\ \neg R$$

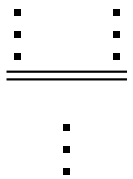
$$\frac{A_0, A_1, M \Rightarrow N}{A_0 \wedge A_1, M \Rightarrow N}\ \wedge L \qquad \frac{M \Rightarrow N, A \qquad M \Rightarrow N, B}{M \Rightarrow N, A \wedge B}\ \wedge R$$

First-order Logical Rules

$$\frac{\forall x A, A(x/t), M \Rightarrow N}{\forall x A, M \Rightarrow N}\ \forall L \qquad \frac{M \Rightarrow N, A(y)}{M \Rightarrow N, \forall x A}\ \forall R$$

where the y does not occur neither in M nor in N.

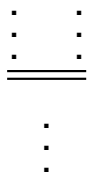
We now add explanatory rules.



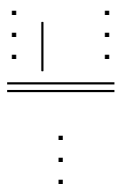
One premise

$$\begin{array}{c} \vdots \\ = \\ \vdots \end{array}$$

Two premises



One premise + one condition



One premise + one condition

$$\dfrac{\vdots \mid \vdots}{\vdots}$$

$$\ldots \mid \ldots \models \ldots$$

Definition

Formulas in contexts, as C in $A[C]$, occur with either a positive or a negative polarity, where polarities are defined in a standard way, e.g. see Troelstra and Schwichtenberg (1996).

Definition

The *converse of a formula* A, written $A^{\perp}$, is defined as follows:

$$A^{\perp} = \begin{cases} \neg^{n-1}E, & \text{if } A = \neg^{n}E \text{ and n is odd} \\ \neg^{n+1}E, & \text{if } A = \neg^{n}E \text{ and n is even} \end{cases}$$

where the main connective in E is not a negation, $n \geqslant 0$ and 0 is taken to be an even number.

Double negation

$$\dfrac{M \Rightarrow N, A[B]}{M \Rightarrow N, A[\neg\neg B]}\ \neg\neg$$

Double negation

$$\frac{M \Rightarrow N, B}{M \Rightarrow N, \neg\neg B}\ \neg\neg$$

Double negation

$$\dfrac{M \Rightarrow N, A[B]}{M \Rightarrow N, A[\neg\neg B]}\ \neg\neg$$

Conjunction/Disjunction

$$\frac{M \Rightarrow N, A[B] \quad M \Rightarrow N, A[C]}{M \Rightarrow N, A[B \circ C]}\ \circ_1 \qquad \frac{M \Rightarrow N, A[B_j] \mid \quad M \Rightarrow N, A[B_i]}{M \Rightarrow N, A[B_1 \circ B_2]}\ \circ_2$$

Conjunction/Disjunction

$$\frac{M \Rightarrow N, A[B] \quad M \Rightarrow N, A[C]}{M \Rightarrow N, A[B \circ C]} \circ_1 \qquad \frac{M \Rightarrow N, A[B_j] \mid \quad M \Rightarrow N, A[B_i]}{M \Rightarrow N, A[B_1 \circ B_2]} \circ_2$$

If $\circ = \wedge^+, \vee^-$ and it is either in the scope of no quantifier or in the scope of $(\forall x)^+$ or $(\exists x)^-$, then $\circ_1$.

Conjunction/Disjunction

$$\frac{M \Rightarrow N, A[B] \quad M \Rightarrow N, A[C]}{M \Rightarrow N, A[B \circ C]} \circ_1 \qquad \frac{M \Rightarrow N, A[B_j] \mid \quad M \Rightarrow N, A[B_i]}{M \Rightarrow N, A[B_1 \circ B_2]} \circ_2$$

If $\circ = \wedge^+, \vee^-$ and it is either in the scope of no quantifier or in the scope of $(\forall x)^+$ or $(\exists x)^-$, then $\circ_1$.
If $\circ = \vee^+, \wedge^-$ and it is either in the scope of no quantifier or in the scope of $(\forall x)^-$ or $(\exists x)^+$, then $\circ_1$ and $\circ_2$.

Conjunction/Disjunction

$$\frac{M \Rightarrow N, A[B] \quad M \Rightarrow N, A[C]}{M \Rightarrow N, A[B \wedge^{+} C]} \circ_1 \qquad \frac{M \Rightarrow N, A[B_j] \mid \quad M \Rightarrow N, A[B_i]}{M \Rightarrow N, A[B_1 \circ B_2]} \circ_2$$

If $\circ = \wedge^{+}, \vee^{-}$ and it is either in the scope of no quantifier or in the scope of $(\forall x)^{+}$ or $(\exists x)^{-}$, then $\circ_1$.
If $\circ = \vee^{+}, \wedge^{-}$ and it is either in the scope of no quantifier or in the scope of $(\forall x)^{-}$ or $(\exists x)^{+}$, then $\circ_1$ and $\circ_2$.

Conjunction/Disjunction

$$\frac{M \Rightarrow N, A[B] \quad M \Rightarrow N, A[C]}{M \Rightarrow N, A[B \circ C]}\ \circ_1 \qquad \frac{M \Rightarrow N, A[B_j] \mid \quad M \Rightarrow N, A[B_i]}{M \Rightarrow N, A[B_1 \vee^+ B_2]}\ \circ_2$$

If $\circ = \wedge^+, \vee^-$ and it is either in the scope of no quantifier or in the scope of $(\forall x)^+$ or $(\exists x)^-$, then $\circ_1$.
If $\circ = \vee^+, \wedge^-$ and it is either in the scope of no quantifier or in the scope of $(\forall x)^-$ or $(\exists x)^+$, then $\circ_1$ and $\circ_2$.

$$\dfrac{\Rightarrow \forall x(Zx \to Nx) \quad \Rightarrow \forall x(SNx \to Nx)}{\Rightarrow \forall x((Zx \vee SNx) \to Nx)}\ \circ 1$$

Negation of Conjunction/Disjunction

$$\frac{M \Rightarrow N, A[B^{\perp}] \quad M \Rightarrow N, A[C^{\perp}]}{M \Rightarrow N, A[\neg(B \circ C)]}\ \neg\circ_1 \qquad \frac{M \Rightarrow N, A[B_j^{\perp}] \mid \quad M \Rightarrow N, A[B_i^{\perp}]}{M \Rightarrow N, A[\neg(B_1 \circ B_2)]}\ \neg\circ_2$$

Negation of Conjunction/Disjunction

$$\frac{M \Rightarrow N, A[B^{\perp}] \quad M \Rightarrow N, A[C^{\perp}]}{M \Rightarrow N, A[\neg(B \circ C)]}\ \neg\circ_1 \qquad \frac{M \Rightarrow N, A[B_j^{\perp}] \mid \quad M \Rightarrow N, A[B_i^{\perp}]}{M \Rightarrow N, A[\neg(B_1 \circ B_2)]}\ \neg\circ_2$$

If $\circ = \vee^{-}, \wedge^{+}$ and it is either in the scope of no quantifier or in the scope of $(\forall x)^{+}$ or $(\exists x)^{-}$, then $\neg\circ_1$.

Negation of Conjunction/Disjunction

$$\frac{M \Rightarrow N, A[B^{\perp}] \quad M \Rightarrow N, A[C^{\perp}]}{M \Rightarrow N, A[\neg(B \circ C)]}\ \neg\circ_1 \qquad \frac{M \Rightarrow N, A[B_j^{\perp}] \mid \quad M \Rightarrow N, A[B_i^{\perp}]}{M \Rightarrow N, A[\neg(B_1 \circ B_2)]}\ \neg\circ_2$$

If $\circ = \vee^{-}, \wedge^{+}$ and it is either in the scope of no quantifier or in the scope of $(\forall x)^{+}$ or $(\exists x)^{-}$, then $\neg\circ_1$.
If $\circ = \wedge^{-}, \vee^{+}$ and it is either in the scope of no quantifier or in the scope of $(\forall x)^{-}$ or $(\exists x)^{+}$, then $\neg\circ_1$ and $\neg\circ_2$.

$$\dfrac{\Rightarrow \neg p \quad \Rightarrow \neg q}{\Rightarrow \neg(p \vee q)}\ \neg\circ 1$$

Quantifiers

$$\frac{M \Rightarrow N, A[By]}{M \Rightarrow N, A[\odot x.Bx]}\ \odot_1 \qquad \frac{M \Rightarrow N, A[\odot x.Bx], A[Bt]}{M \Rightarrow N, A[\odot x.Bx], A[\odot x.Bx]}\ \odot_2$$

where y does not occur in M nor in N.

Quantifiers

$$\frac{M \Rightarrow N, A[By]}{M \Rightarrow N, A[\odot x.Bx]}\ \odot_1 \qquad \frac{M \Rightarrow N, A[\odot x.Bx], A[Bt]}{M \Rightarrow N, A[\odot x.Bx], A[\odot x.Bx]}\ \odot_2$$

where y does not occur in M nor in N.

If $\odot = \forall^+, \exists^-$ and it is either in the scope of no other quantifier or in the scope of $(\forall x)^+$ or $(\exists x)^-$, then $\odot 1$.

Quantifiers

$$\frac{M \Rightarrow N, A[By]}{M \Rightarrow N, A[\odot x.Bx]}\ \odot_1 \qquad \frac{M \Rightarrow N, A[\odot x.Bx], A[Bt]}{M \Rightarrow N, A[\odot x.Bx], A[\odot x.Bx]}\ \odot_2$$

where y does not occur in M nor in N.

If $\odot = \forall^+, \exists^-$ and it is either in the scope of no other quantifier or in the scope of $(\forall x)^+$ or $(\exists x)^-$, then $\odot 1$.
If $\odot = \exists^+, \forall^-$, then $\odot 2$.

Negation of Quantifiers

$$\frac{M \Rightarrow N, A[B^{\perp} y]}{M \Rightarrow N, A[\neg(\odot x.Bx)]}\ \odot_1 \qquad \frac{M \Rightarrow N, A[\neg(\odot x.Bx)], A[B^{\perp} t]}{M \Rightarrow N, A[\neg(\odot x.Bx)], A[\neg(\odot x.Bx)]}\ \odot_2$$

where y does not occur in M nor in N.

Negation of Quantifiers

$$\frac{M \Rightarrow N, A[B^{\perp} y]}{M \Rightarrow N, A[\neg(\odot x.Bx)]}\ \odot_1 \qquad \frac{M \Rightarrow N, A[\neg(\odot x.Bx)], A[B^{\perp} t]}{M \Rightarrow N, A[\neg(\odot x.Bx)], A[\neg(\odot x.Bx)]}\ \odot_2$$

where y does not occur in M nor in N.

If $\odot = \exists^{-}, \forall^{+}$ and it is either in the scope of no quantifier or in the scope of $(\forall x)^{+}$ or $(\exists x)^{-}$, then $\odot 1$.

Negation of Quantifiers

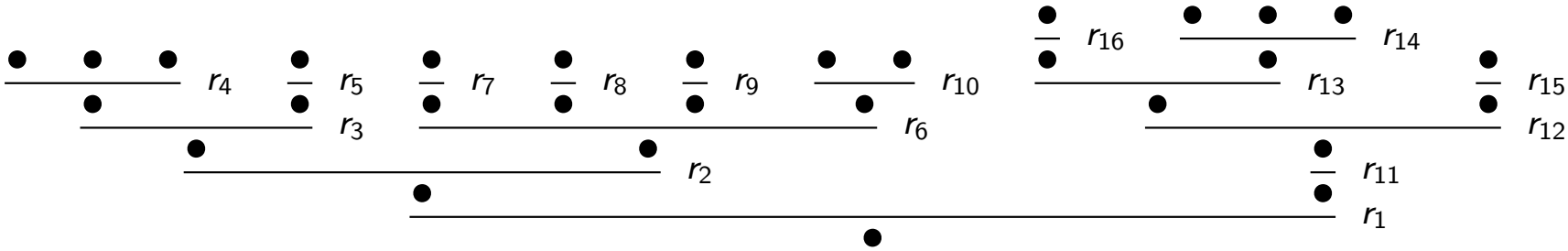
$$\frac{M \Rightarrow N, A[B^{\perp}y]}{M \Rightarrow N, A[\neg(\odot x.Bx)]}\ \odot_1 \qquad \frac{M \Rightarrow N, A[\neg(\odot x.Bx)], A[B^{\perp}t]}{M \Rightarrow N, A[\neg(\odot x.Bx)], A[\neg(\odot x.Bx)]}\ \odot_2$$

where y does not occur in M nor in N.

If $\odot = \exists^{-}, \forall^{+}$ and it is either in the scope of no quantifier or in the scope of $(\forall x)^{+}$ or $(\exists x)^{-}$, then $\odot 1$.
If $\odot = \forall^{-}, \exists^{+}$, then $\odot 2$.

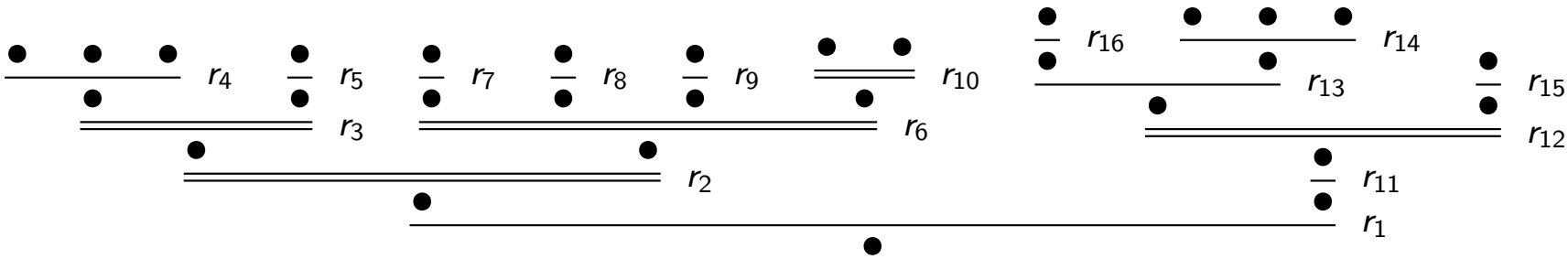
Thanks to the explanatory rules, we can now construct:

Thanks to the explanatory rules, we can now construct:



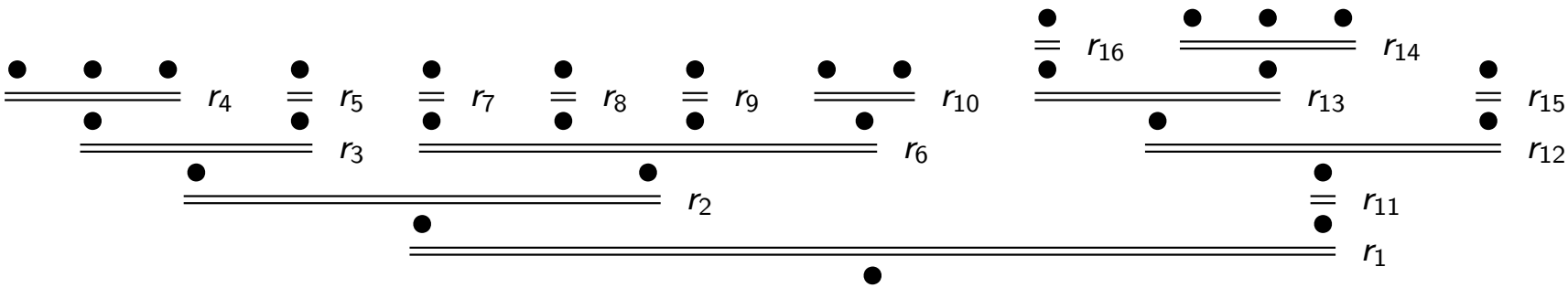
Derivations

Thanks to the explanatory rules, we can now construct:



Derivations with explanatory steps

Thanks to the explanatory rules, we can now construct:



Formal explanations

Definition

A derivation in **Gfcl** is a finite (upwardgrowing) tree with a single root. The nodes of the tree are labelled by sequents or sequents with a bar and the top nodes are labelled by initial sequents. For each non-terminal node, its label is connected with the labels of the immediate predecessor nodes according with one of the logical rules or one of the explanatory rules. The root of the tree is the conclusion of the whole derivation and its label is a theorem of the sequent calculus, in symbol $\vdash_{Gfcl} M \Rightarrow N$.

Definition

For any multiset of sequents S' (which might be empty), and for any multiset of sequents S, we say that under the condition $(S')^{\perp}$, there exists a complete and immediate formal explanation from S to $M \Rightarrow N$, in symbols $S' \mid S \Vdash M \Rightarrow N$ if, and only if, one of the explanatory rules $\neg\neg, \circ_1, \circ_2, \neg\circ_1, \neg\circ_2$ link S', S and $M \Rightarrow N$.

Definition

For any multiset of sequents S' (which might be empty), and for any multiset of sequents S, we say that under the condition $(S')^{\perp}$, S completely and mediately formally explain $M \Rightarrow N$, in symbols $S' \mid S \Vdash^{m} M \Rightarrow N$ if, and only if:

- $S' \mid S \Vdash M \Rightarrow N$,
- $S'' \mid S''' \Vdash M' \Rightarrow N'$ $S'''' \mid S''''', M' \Rightarrow N' \Vdash M \Rightarrow N$, and $S'' \cup S'''' = S'$ and $S''' \cup S''''' = S$.

Results/Directions of Future Research

Theorem (Soundness)

For any mutisetset of sequents (possibly empty) S', S, and sequent $M \Rightarrow N$,

$$\textit{if } S' \mid S \Vdash M \Rightarrow N, \textit{ then } (S')^{\tau} \mid (S)^{\tau} \Vdash \bigwedge M \rightarrow \bigvee N$$

where $(S')^{\tau}$, $(S)^{\tau}$ are the mulistets of sequents standardly translated into formulas.

Theorem (Completeness)

For any mutisetset of closed formulas (possibly empty) N', N, and formula $A[C]$,

$$\textit{if } N' \mid N \Vdash A[C], \textit{ then } (N')^{\delta} \mid (N)^{\delta} \Vdash \Rightarrow A[C]$$

where for any $M = \{A[B_1], ...A[B_n]\}$, $(M)^{\delta} = \{\Rightarrow A[B_1], ..., \Rightarrow A[B_n]\}$.

Theorem (Admissibility)

Any explanatory rule is admissible in **Gfcl**.

Application

We use formal explanations to properly modeling mathematical explanations.[a]

[a] See *Mathematical explanations: an analysis via complexity and proof*, Submitted, 2023. Also in a joint work with E. Pimentel.

Directions of future research

- Formal explanation/formal grounding for logics different from classical first-order logic (intuitionistic logic, modal logic, ...).
- Explanatory rules and proof-theoretic semantics.[3]
- Explanatory rules and deep inferences.[4]
- Conceptual explanation/grounding and the links with causal explanation/causality.
- Formal explanations and explainable AI.

[3] F. Poggiolesi, Grounding rules and (hyper-)isomorphic formulas, *Australasian Journal of Logic*, 17: 70-80, 2020.

[4] See http://alessio.guglielmi.name/res/cos/

THANK YOU!