Notes for the Logic and Metaphysics course

October 27, 2018

1 Propositional modal logic

1.1 Classical Propositional Logic: a Brief Remainder

Language In these notes V will denote a fixed infinite set, whose all elements can be arranged in an infinite sequence whose positions are given by natural numbers (such sets are called *countable*; there are sets that do not have this property, e.g. the set of real numbers. They are called *uncountable*). Elements of V are called *propositional variables* and are denoted p, p_1, $p_2, \ldots$, q, $q_1, \ldots, r, r_1, \ldots$.

Definition 1 (Formulae of CPL)**.** The set of formulae F_{prop} of CPL is the least set P such that

1. $V \subseteq F_{prop}$
2. if ϕ, ψ are any elements of F_{prop}, then so are $(\phi \wedge \psi)$, $(\phi \vee \psi)$, $(\neg\phi)$ and $(\phi \to \psi)$.

The clause "F_{prop} is the least set P such that conditions 1. and 2. hold" means that F_{prop} satisfies 1 and 2 and it is a subset of every set which satisfies 1. and 2.

Semantics Now we proceed to the semantics of CPL. Definition is surprisingly simple and we will explain its meaning in a moment.

Definition 2 (Models for CPL)**.** A model M for CPL is a subset of V.

Interpretation: Definition becomes more intuitive if we think of elements of V as atomic facts that can hold independently of each other. A model is then a specification which atomic facts hold (i.e. elements of M) and which do not (i.e. those left outside M).

Let us note that to define a subset of V is *essentially* the same as to define a function f mapping each element of V either to 0 or 1 (the choice of 0 and 1 is purely conventional - any two distinct object will do). Indeed, if a subset M of V is given then function f can be defined as sending the elements of M to 1 and the elements not belonging to M to 0. In the other direction, if f is given, then a subset of V can be defined as the set of all those elements which are sent by f to 1. Sometimes we will make use of this double characterization, treating a model M as a function, which we will call a *valuation*.

Definition 3 (Satisfaction relation)**.** Let M be a model for CPL and ϕ be a formula of CPL. We shall say that $M \models \phi$ iff

1. $\phi = p$ for some $p \in V$ and $p \in M$, or
2. $\phi = \psi_1 \wedge \psi_2$ and $M \models \psi_1$ and $M \models \psi_2$, or
3. $\phi = \psi_1 \vee \psi_2$ and $M \models \psi_1$ or $M \models \psi_2$, or
4. $\phi = \psi_1 \to \psi_2$ and if $M \models \psi_1$, then $M \models \psi_2$, or
5. $\phi = \neg\psi$ and it is not the case that $M \models \psi$.

Example 4. Let $M = \{p, q\}$ then

$$M \models p \wedge ((p \to r) \to q)$$

since:

1. $M \models p$
2. $M \not\models r$
3. $M \not\models p \to r$ (because of the two above observations)
4. $M \models ((p \to r) \to q)$ (because the antecedent is false)

Proof System

Definition 5 (Hilbert-Style Proof System)**.** *Axiom* is any sentence of the form

1. $\neg\phi \to (\phi \to \psi)$
2. $\phi \to (\psi \to \phi)$

3. $(\phi \to \psi) \to ((\neg\phi \to \psi) \to \psi)$

4. $(\phi \to (\psi \to \theta)) \to ((\phi \to \psi) \to (\phi \to \theta))$

more precisely: any formula resulting from substituting elements of F_{Prop} for ϕ, ψ, θ in one of the above is our axiom. For example

$$(\neg(p \to p)) \to ((p \to p) \to (p \wedge q))$$

is an axiom because it results from the substitution $\phi \mapsto (p \to p)$, $\psi \mapsto (p \wedge q)$. We adopt also a single rule of reasoning - Modus Ponens

$$\frac{\phi, \phi \to \psi}{\psi}$$

A proof of ϕ is any sequence

$$\psi_1, \psi_2, \ldots, \psi_n$$

of elements of F_{Prop} such that

1. each of $\psi_1, \ldots, \psi_n$ is either an axiom or can be obtained via Modus Ponens rule from previous elements in our sequence;

2. $\psi_n = \phi$.

Sentence ϕ *is provable* if and only if there exists a proof of ϕ. We shall denote the symbol

$$\vdash \phi$$

to denote that ϕ is provable.

It can be checked that Hilbert-Style Proof System as defined above matches the natural semantics that we gave earlier. More precisely it holds that:

Theorem 6 (Completeness theorem for Hilbert-Style Calculus)**.** *For every* ϕ *from* F_{prop}

$\vdash \phi$ if and only if for every model M, $M \models \phi$

The left-to-right part of the above theorem states that the proof system we defined is *sound* - whatever can be deduced in it is true in every "possible world" (i.e. every model for our language). The right-to-left direction states that it is *complete*: whatever is true in every possible world can be justified by reasoning formalized in this calculus.

Example 7. The calculus we defined is not very convenient to work in. For example the following sequence of formulae is a proof of $p \to p$:

1. $(p \to ((p \to p) \to p)) \to ((p \to (p \to p)) \to (p \to p))$ (instantiation of axiom 4 with $\phi \mapsto p$, $\psi \mapsto (p \to p)$, $\theta \mapsto p$)
2. $(p \to ((p \to p) \to p))$ (instantiation of axiom 2 with $\phi \mapsto p$, $\psi \mapsto (\to p)$)
3. $((p \to (p \to p)) \to (p \to p))$ (application of Modus Ponens to the two formulae above)
4. $(p \to (p \to p))$ (insantiation of axiom 2 with $\phi \mapsto p$, $\psi \mapsto p$)
5. $(p \to p)$ (application of Modus Ponens to the two above formulae.)

However it is often used because of the very short definition and some additional theoretical features. We will give a more convenient decision procedure (algorithm) for deciding whether a formula is valid.

Let us observe that if we have two models $M \subseteq V, N \subseteq V$ that *agree* on set of variables $\{p_0, \ldots, p_n\}$, i.e. for each i, p_i belongs to M if and only if p_i belongs to N, then for every formula ϕ which uses only $\{p_0, \ldots, p_n\}$ as propositional variables

$$M \models \phi \text{ if and only if } N \models \phi$$

For example if $M = \{p_0, p_1, p_2\}$, $N = \{p_0, p_1, q_0\}$ then these models satisfy the same formulae with propositional variables p_0, p_1, p_3. For example

$$M \models (p_0 \to \neg p_1) \wedge \neg p_3 \text{ if and only if} N \models (p_0 \to \neg p_1) \wedge \neg p_3$$

It follows that to check whether a formula ϕ is satisfied in a model M we need only finitely many informations about M: for each propositional variable which occurs in ϕ (there are finitely many of them!) we have to know whether it belongs to M or not. Hence it is enough to check each of finitely many possible cases. For example let $\phi = (p \wedge q) \to r$. Then it is sufficient to check what happens if

1. all three p, q, r belong to our model.
2. from p, q, r only p belongs to our model. (similarly for q, r)
3. from p, q, r only p, q belong to our model (similarly for (q, r), (p, r))
4. none of p, q, r belongs to our model.

Let us note that this corresponds to considering all possible assignmets of $0, 1$ (0 for "does not hold", 1 for "holds") to $\{p, q, r\}$. This gives rise to the method of "truth tables": to check whether formula is valid (provable in our Proof System) it is enough to check whether every valuation of its propositional variables satisfies it.

Example 8. Let us check that one of axioms of the Hilbert-Style Proof System is valid by checking all valuations. Let us consider

$$\phi = p \to (q \to p)$$

ϕ contains only two propositional variables p, q. There are four valuations that we have to consider

1. $\begin{cases} p \mapsto 1 \\ q \mapsto 1 \end{cases}$
   (this corresponds to the situation in which model contains both p and q)
2. $\begin{cases} p \mapsto 1 \\ q \mapsto 0 \end{cases}$
   (this corresponds to the situation in which our model contains p but not q)
3. $\begin{cases} p \mapsto 0 \\ q \mapsto 1 \end{cases}$
   (this corresponds to the situation in which a model contains q but not p)
4. $\begin{cases} p \mapsto 0 \\ q \mapsto 0 \end{cases}$
   (this corresponds to the situation in which our model contains neither p nor q)

It is easy to verify that each valuation which sends p to 1 makes ϕ true, since it makes true the implication in the succedent of ϕ. Moreover each valuation which sends p to 0 makes ϕ true since it makes the antecedent of ϕ false. Hence every valuation makes ϕ true, hence ϕ is valid.

The following formula

$$\psi = (p \vee q) \to p \wedge q$$

is not valid since the following valuation

$$\begin{cases} p \mapsto 1 \\ q \mapsto 0 \end{cases}$$

makes ψ false.

In the next exercise $p \leftrightarrow q$ abbreviates $(p \to q) \wedge (q \to p)$.

Exercise 1. Check whether the following formulae are valid. In case they are not valid (i.e. there exists a falsifying valuation) check whether there exists a valuation which makes them true.

1. $(p_0 \to p_0) \to ((\neg p_0) \to p_0)$
2. $(p_0 \to \neg p_0) \to q_0$
3. $((p_0 \vee p_1) \vee p_2) \vee (p_3 \vee (p_4 \vee \neg p_4))$
4. $((p_0 \vee p_2) \wedge \neg p_2) \to p_3$
5. $(p \to q) \to ((p \wedge r) \to q)$
6. $(p \leftrightarrow q) \vee ((q \leftrightarrow r) \vee (p \leftrightarrow r))$
7. $(p \leftrightarrow q) \vee (q \leftrightarrow r)$.
8. $((p \wedge q) \to r) \to ((p \to r) \wedge (q \to r))$
9. $((p \vee q) \to r) \to ((p \to r) \vee (q \to r))$
10. $((p \wedge q) \to r) \to ((p \to r) \vee (q \to r))$

1.2 The syntax of the propositional modal logic

Propositional modal logic is an extension of the classical propositional logic in which to the connectives $\neg, \wedge, \vee, \to$ add two unary (i.e. syntactically behaving like a negation) operators $\Box, \Diamond$ with the intended reading "it is necessary that..." and "it is possible that..." respectively.

Formally, the set of propositional modal formulae over the set of propositional variables V is defined as the smallest set F with the following properties:

1. Any element of V belongs to F.
2. If $\phi \in F$, then $\neg\phi \in F$.
3. If both $\phi \in F$ and $\psi \in F$, then $\phi \wedge \psi \in F$, $\phi \vee \psi \in F$ and $\phi \to \psi \in F$.
4. If $\phi \in F$, then $\Box\phi \in F, \Diamond\phi \in F$.

Slightly unwinding the formal definition, we see that propositional modal formulae look like the usual propositional formulae, but we allow to write additional symbols $\Box, \Diamond$ in front of formulae, like negations. We will write the elements of V with the letters $p, q, r, s, \ldots$ or sometimes with subscripts $p_1, p_2, \ldots, q_1, q_2, \ldots$. So as an example of the modal formula we have:

$$\Box(\Diamond p \to \neg\Box\neg q)$$

or

$$\Box(q \vee \Box\Diamond\Box r) \vee (p_1 \to \Diamond\Box\Box p_3).$$

1.3 Kripke models

We will now describe the semantics of the propositional modal logic.

Definition 9. By a **Kripke model** we mean a tuple $\langle K, R, f\rangle$, where

1. K is an arbitrary nonempty set.
2. $R \subset K^2$ is an arbitrary relation.
3. f is an arbitrary function with the domain V, which assigns to every letter $p \in V$ a subset of K.

The above definition is probably somewhat unenlightening, so let's try to elaborate on it. Typically, when we define something as a tuple, we think of such an object as a set with some additional structures: relations, functions etc. defined on this set. In our case, we usually call the elements of K **the possible worlds**, R is called the **accessibility relation** and f is called the **valuation function**.

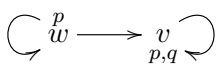
We should think of Kripke models as of sets of worlds for which the relation between w and v holds if and only if v is possible from the point of view of w. For each of these worlds w, we explicitly indicate which atomic facts hold in these worlds — this is what the valuation function does. If a world w belongs to $f(p)$ this intuitively means that the atomic fact p holds in the world w.

We capture this intuition, when relating Kripke models to propositional modal formulae.

Definition 10. Let $\mathcal{K} = (K, R, f)$ be an a arbitrary Kripke model and w a world in this model. We define what does it mean for a modal formula ϕ to be satisfied in the model $\mathcal{K}$ in the world w (which we write $\mathcal{K}, w \models \phi$) by induction on complexity of ϕ.

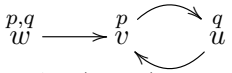
1. If p is a propositional variable, then $\mathcal{K}, w \models p$ iff $w \in f(p)$.
2. $\mathcal{K}, w \models \neg\phi$ iff it is not the case that $\mathcal{K}, w \models \phi$.
3. $\mathcal{K}, w \models \phi \wedge \psi$ iff $\mathcal{K}, w \models \phi$ and $\mathcal{K}, w \models \psi$.
4. $\mathcal{K}, w \models \phi \vee \psi$ iff $\mathcal{K}, w \models \phi$ or $\mathcal{K}, w \models \psi$.
5. $\mathcal{K}, w \models \phi \to \psi$ iff it is not the case that $\mathcal{K}, w \models \phi$ or $\mathcal{K}, w \models \psi$.
6. $\mathcal{K}, w \models \Box\phi$ iff $\mathcal{K}, v \models \phi$ for all v such that $R(w, v)$.
7. $\mathcal{K}, w \models \Diamond\phi$ iff $\mathcal{K}, v \models \phi$ for some v such that $R(w, v)$.

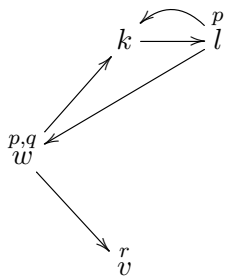
Thus for example in a model $\mathcal{K} = (W, R, f)$ with two worlds w, v such that the relation R holds only between pairs $\langle w, w\rangle, \langle w, v\rangle, \langle v, v\rangle$ and the valuation function f which assigns $\{w, v\}$ to p and $\{v\}$ to q we have for example: $\mathcal{K}, w \models \neg\Box q \wedge \Box p \wedge \Box\Diamond q$. This model can be depicted



Let us verify that $\mathcal{K}, w \models \Box\Diamond q$, the rest of cases being rather easy. Unfolding the definition we get that $\mathcal{K}, w \models \Box\Diamond q$ if and only if for every world w' which is in relation R with w there exists a possible world w'' in relation R with w' such that w'' satisfy q. This is true in the model given above since only w and v are in the relation with w and each of them sees world v which satisfies q.

Exercise 2. Check whether given formulae hold in the given model at world w:

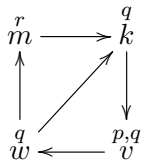
1. 

formulae: $\Box p$, $\Diamond\Diamond p$, $\Diamond\Diamond\Diamond p$, $\Diamond\Box(p \vee q)$.
2. 

(w, k, l, v are worlds and p, q, r are propositional variables). Formulae:

(a) $\Box\Diamond p$,
(b) $\Diamond\Box p$,
(c) $q \to (\Diamond\neg p \to \Diamond\Diamond p)$,
(d) $\Diamond r \to \Box\neg q$

3.

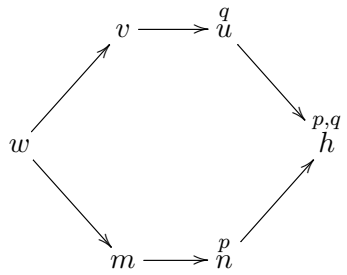


Formulae:

(a) $\Box r$
(b) $\Box\Diamond q$
(c) $\Diamond q \to \Box r$
(d) $\Diamond q \to \Box\Diamond q$

Exercise 3. Check whether given formulae hold in the given model at world w:

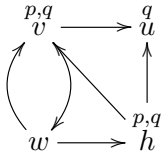
1.



formulae:

(a) $\Diamond\Box p \to \Box\Diamond p$
(b) $\Diamond\Diamond p \to \Diamond\Diamond q$

2.



(a) $\Diamond\Diamond\Diamond\Diamond q \to \Diamond\Diamond\Diamond q$
(b) $\Diamond\Diamond\Diamond\Box q \to \Box\Diamond\Diamond q$

Exercise 4. Give an example of Kripke structures M_i and worlds w_i such that:

1. $M_1, w_1 \models \Box(p \vee q) \wedge \neg(\Box p \vee \Box q)$.
2. $M_2, w_2 \nVdash p \to \Diamond p$.
3. $M_3, w_3 \models \Box p \wedge \neg\Box\Box p$.
4. $M_4, w_4 \models \Diamond p \wedge \neg\Box\Diamond p$.
5. $M_5, w_5 \models \Box(p \to q) \to (\Box p \to \Box q)$
6. $M_6, w_6 \models \Diamond\Diamond\Diamond\Diamond p \wedge (\neg\Diamond p \wedge (\neg\Diamond\Diamond p \wedge \neg\Diamond\Diamond\Diamond p))$
7. $M_7, w_7 \Vdash \Diamond\Diamond\Diamond\Box(p \wedge \neg p)$
8. $M_8, w_8 \nVdash \Box\Box p \to \Box p$

Exercise 5. Check whether the following formulae are tautologies of Modal Logic, i.e. whether they hold in all Kripke models.

1. $\Box\Diamond p \to \Diamond\Diamond p$
2. $(\Box\Diamond p \wedge \Diamond q) \to \Diamond\Diamond p$
3. $(\Diamond\Box p \wedge \Diamond\Diamond q) \to \Diamond\Diamond p$
4. $\Box(p \wedge q) \to \Box p$
5. $\Box(p \wedge q) \to (\Box p \wedge \Box q)$
6. $\Diamond(p \vee q) \to \Diamond p \vee \Diamond q$
7. $(\Diamond p \wedge \Diamond q) \to \Diamond(p \wedge q)$
8. $(\Box p \equiv \Box q) \to (\Diamond p \equiv \Diamond q)$
9. $\Diamond p \to \Box\Diamond p$
10. $\Box(\Box p \to p) \to \Box p$
11. $\Box(\Diamond p \vee \Diamond\neg p)$.
12. $\Box\big((\Diamond p \wedge \Box(p \to q)) \to \Diamond q\big)$
13. $\Diamond\big((\Diamond p \wedge \Box(p \to q)) \to \Diamond q\big)$
14. $\Box\big((p \wedge (p \to q)) \to \Diamond q\big)$

1.4 Proof system and the most important theories

The next definition introduces a Hilbert-style proof system for modal logic. It is a proper extension of the proof system defined for Classical Propositional Calculus.

Definition 11 (Hilbert-style proof system for ML)**.** Hilbert-style proof system for ML contains as axioms

1. all instantiations of tautologies of Classical Propositional Calculus with formulae of modal logic.
2. all instantiations of the following scheme

$$\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$$

   with formulae of modal logic.

We have two rules of reasoning:

1. Modus Ponens (the same as for Classical Propositional Calculus)
2. Gödels Rule, or Necessitation

$$\frac{\phi}{\Box\phi}$$

Definition of a proof and provability is the same as previously.

Example 12. If ϕ is an instantiation of a tautology of Classical Propositional Calculus, then $\Box\phi$ (and $\Box\Box\phi$, $\Box\Box\Box\phi$...) is provable. For example a proof of $\Box(p \to p)$ is the following sequence of length two

$$p \to p, \Box(p \to p).$$

The above proof system "matches" the semantics for modal logic we introduced earlier - more precisely we have the following theorem. Before stating it it will be convenient to introduce one more definition:

Definition 13. Let $\mathcal{K} = \langle K, R, f \rangle$ be a Kripke model and ϕ a formla of modal logic. We shall say that ϕ is true in $\mathcal{K}$ and write

$$\mathcal{K} \models \phi$$

if for every world $w \in K$ we have

$$\mathcal{K}, w \models \phi$$

Theorem 14 (Completeness Theorem for modal logic)**.** *For every formula ϕ of modal logic*

$$\phi \text{ is provable if and only if for every Kripke model } \mathcal{K}, \mathcal{K} \models \phi$$

This theorem can be strengthened. Let us introduce one natural generalization of the standard provability relation.

Definition 15. Let Th be a set of modal formulae and ϕ a modal formula. We shall say that ϕ *is a consequence of* Th (or that Th proves ϕ) and write $\text{Th} \vdash \phi$ if and only if there is a proof of ϕ in which formulae from Th can occur as axioms.

We say that a Kripke model $\mathcal{K}$ satisfies a set of formulae Th (and write it $\mathcal{K} \models \text{Th}$) if it satisfies every formula from Th.

Theorem 16 (Completeness Theorem for modal logic, 2)**.** *For every formula ϕ of modal logic and every set of modal formulae* Th*,*

$$\text{Th} \vdash \phi \text{ if and only if for every Kripke model } \mathcal{K}, \text{if } \mathcal{K} \models \text{Th}, \text{ then } \mathcal{K} \models \phi$$

Moreover we have the following very useful theorem (analogous to the one known from the First Order Logic)

Theorem 17 (Deduction Theorem)**.** *Let ϕ, ψ be modal formulae and* Th *a set of modal formulae. We have that*

$$\text{Th} \vdash \phi \to \psi \text{ if and only if } \text{Th} \cup \{\phi\} \vdash \psi$$

Modal logic introduced so far is a very general tool and can be adapted to model different notions. Proof calculus gives us an easy way to adapt the current formalism to new situations - we might simply add new axioms and investigate into so formed systems. Let us give some examples of systems of modal logic which can be found in the literature:

Definition 18.

1. System K contains all the sentences provable in the Hilbert-style proof calculus for ML.
2. System T contains all the sentences provable in the Hilbert-style proof calculus for ML in which all instantiations of the scheme

$$\Box\phi \to \phi \tag{T}$$

can be taken as axioms.

3. System $S4$ contains all the sentences provable in the Hilbert-style proof calculus for ML in which all instantiations of the schemes

$$\Box\phi \to \phi \tag{T}$$

$$\Box\phi \to \Box\Box\phi \tag{4}$$

can be taken as axioms.

4. System $S5$ contains all the sentences provable in the Hilbert-style proof calculus for ML in which all instantiations of the schemes

$$\Box\phi \to \phi \tag{T}$$

$$\Box\phi \to \Box\Box\phi \tag{4}$$

$$\Diamond\phi \to \Box\Diamond\phi \tag{5}$$

can be taken as axioms.

5. System D contains all the sentences provable in the Hilbert-style proof calculus for ML in which all instantiations of the scheme

$$\Box\phi \to \Diamond\phi \tag{D}$$

can be taken as axioms.

Example 19. Every sentence of the form

$$\phi \to \Box\Diamond\phi \tag{B}$$

is in $S5$ (i.e. it is provable from axiom schemata (T),(4) and (5)). Indeed, let us fix any ϕ. We shall show that $S5 \cup \{\phi\} \vdash \Box\Diamond\phi$ which clearly suffices by Deduction Theorem (Theorem 17). We have $\phi \to \Diamond\phi$ by the contraposition of (T) axiom for $\neg\phi$. Hence by Modus Ponens we have $\Diamond\phi$. Using the instantiation of (5) axiom schema for ϕ we get that $\Diamond\phi \to \Box\Diamond\phi$. Hence by Modus Ponens again we get $\Box\Diamond\phi$, as wanted.

Exercise 6. Show that for arbitrary Kripke model $\mathcal{K} = \langle K, R, f\rangle$ such that R is symmetric[1] and arbitrary formula $\psi = \phi \to \Box\Diamond\phi$ we have

$$\mathcal{K} \models \psi$$

Exercise 7. Show that for arbitrary Kripke model $\mathcal{K} = \langle K, R, f\rangle$ such that R satisfies: for every $w \in K$ there exists w' such that wRw' and arbitrary formula $\psi = \Box\phi \to \Diamond\phi$ we have

$$\mathcal{K} \models \psi$$

[1] i.e. if a world w sees a world w' then w' sees w too. Formally: for every w, w', if wRw', then $w'Rw$.

1.5 Kripke Frames and connections with basic systems of Modal Logic

In the previous section we have seen that for some classes of formulae it is possible to determine whether they hold in a given Kripke model basing on appropriate properties of the accessibility relation only. In this paragraph we shall demonstrate a converse to this phenomenon: we will prove that if certain formulae holds in a Kripke model *regardless of the chosen valuation*, then the accessibility relation has appropriate property. The following definition is a translation of the sentence "formula ϕ holds in a Kripke model regardless of the chosen valuation" to a formal language.

Definition 20. A *Kripke Frame* is a nonempty set K (the universe) and a relation $R \subseteq K^2$. In other words: it is a Kripke model with no valuation function. A formula ϕ is *satisfied* in a Kripke frame (K, R) if and only if for every valuation $f : V \to \mathcal{P}(K)$ (recall that V is a set of propositional variables) we have

$$(K, R, f) \models \phi$$

Abusing the notation a little bit we will use symbol $\models$ for satisfiability in a frame.

Example 21. Let $K = \{a, b\}$ and $R = \{\langle a, a\rangle, \langle a, b\rangle, \langle b, a\rangle\langle b, b\rangle\}$ (hence $R = K^2$; such a graph is also called a clique). Then the formula $\Box p \to \Box\Box p$ is satisfied in (K, R), because we have already showed that it is satisfied in every Kripke model in which the relation is transitive. The formula $\Diamond p$ is not satisfied in (K, R) because for a valuation $f(q) = \emptyset$ for every propositional letter we have

$$(K, R, f), a \not\models \Diamond\phi$$

(to give a counterexample for a satisfiability on a frame it is sufficient to find one valuation and one world such that the resulting Kripke model falsify the formula at a chosen world).

Example 22. The formula $\Box p$ is satisfied in a Kripke Frame $K = \{a, b, c, d\}$ and $R = \emptyset$ because for every valuation f, (K, R, f) is a Kripke model in which every world satisfies $\Box\phi$ for every formula ϕ.

Let K be as previously and $R = \{\langle a, b\rangle, \langle a, a\rangle, \langle c, d\rangle\}$. Then the formula $\Box p \to p$ is not satisfied in (K, R) since for f such that

$$f(p) = \{d\}$$

and $f(p) = \emptyset$ for the rest of propositional variables we have

$$(K, R, f), c \not\models \Box p \to p$$

Generalising the (second part of the) above example we might show the following

Proposition 23. *A frame (K, R) satisfies $\Box p \to p$ if and only if R is reflexive, i.e. for every $k \in K$ it holds that kRk.*

Proof. To be added. □

It turns out that we can find similar properties of the accessibility relation for the rest of additional modal properties we introduced in Definition 18. We will summarize them in the following proposition. By saying that the principle X holds in F we mean that every instantiation of the respective scheme holds at F.

Proposition 24. *Let $F = (V, R)$ be a Kripke frame*

1. *The principle* (T) *holds in F iff R is reflexive, i.e. for any $w \in W$ we have $R(w, w)$.*
2. *The principle* (4) *holds in F iff R is transitive, i.e. for any $w_1, w_2, w_3 \in W$ if $R(w_1, w_2)$ and $R(w_2, w_3)$, then $R(w_1, w_3)$.*
3. *The principle* (B) *holds in F iff R is symmetric, i.e. for any $w, w' \in W$ if $R(w, w')$, then $R(w', w)$ as well.*
4. *The principle* (5) *holds in F iff R is Euclidean, i.e. for any w_1, w_2, w_3 if $R(w_1, w_2)$ and $R(w_1, w_3)$, then $R(w_2, w_3)$.*
5. *The principles* (T),(4),(5) *hold jointly in F iff R is an equivalence relation, i.e. it is symmetric, reflexive and transitive.*

Exercise 8. Prove the above proposition.

Putting the above proposition with the second version of Completeness Theorem (Theorem 16) we get the following Completeness Theorem for the modal systems introduced above.

Theorem 25. *Let ϕ be a modal formula.*

1. *T proves ϕ if and only if for every Kripke model $\mathcal{K} = (K, R, f)$ such that R is reflexive we have $\mathcal{K} \models \phi$.*
2. *B proves ϕ if and only if for every Kripke model $\mathcal{K} = (K, R, f)$ such that R is symmetric we have $\mathcal{K} \models \phi$.*

3. *$S4$ proves ϕ if and only if for every Kripke model $\mathcal{K} = (K, R, f)$ such that R is reflexive and transitive we have $\mathcal{K} \models \phi$.*

4. *$S5$ proves ϕ if and only if for every Kripke model $\mathcal{K} = (K, R, f)$ such that R is an equivalence relation we have $\mathcal{K} \models \phi$.*

1.6 Tableaux for propositional modal logic

In the previous subsection we have defined semantics for the propositional modal logic. We can also introduce this logic via a proof system, i.e. by syntactically defining what constitutes a correct proof of a formula. In our case, the proof system to be defined will have one more feature: for any provable formula we will be able to algorithmically *find* its proof, not only *check* whether something is a proof. Let us first introduce a few (very standard) preparatory notions:

Definition 26. By a (directed) **graph** we mean a pair (V, E), where V is any nonempty set and E is any binary relation on V. A **tree** is a graph in which there is no element v such that for some two elements v_1, v_2 both $E(v_1, v)$ and $E(v_2, v)$ holds and there is exactly one element v such that for no v' $E(v, v')$. A **chain** is a subset $\{v_1, \ldots, v_n\} \subseteq V$ such that $E(v_1, v_2)$, $E(v_2, v_3)$, $E(v_3, v_4), \ldots, E(v_{n-1}, v_n)$. A chain in a tree is called **a branch** if no element can be added to it so that it remains a chain (so it is a maximal chain). By a **labelled** graph with labels from the set A we mean a pair (V, E, f), where (V, E) is a graph and f is a function whose domain is V and whose values are elements of A.

Let us elaborate a bit on the above definitions. When talking about pairs (V, E) we really simply think of a set V with some additional structure: in our case E means that we join some elements of V with arrows. So: we have a set with a couple of arrows from one element to the other. We often call elements of graphs **vertices** and pairs in E **edges**. When graph is labelled we simply put another layer of structure on a graph: on some vertices we write some notes containing extra information.

Intuitively a tree is simply a graph, where from one vertex more than one arrow can go, but no two arrows meet together again in one element. On top of that we require, that there is only one vertex which is not pointed at by any arrow. We call it the **root** of the tree.

Let us define first **de Morgan** form ϕ^M of a propositional modal formula ϕ by induction on the shape of ϕ:

1. $p^M = p$ and $(\neg p)^M = \neg p$ for all propositional variables p.

2. $(\phi \wedge \psi)^M = \phi^M \wedge \psi^M$.
3. $(\phi \vee \psi)^M = \phi^M \vee \psi^M$.
4. $(\Box\phi)^M = \Box\phi^M$.
5. $(\Diamond\phi)^M = \Diamond\phi^M$.
6. $(\phi \to \psi)^M = (\neg\phi)^M \vee \psi^M$.
7. $(\neg(\phi \wedge \psi))^M = (\neg\phi)^M \vee (\neg\psi)^M$.
8. $(\neg(\phi \vee \psi))^M = (\neg\phi)^M \wedge (\neg\psi)^M$.
9. $(\neg\neg\phi)^M = \phi^M$.
10. $(\neg\Box\phi)^M = \Diamond(\neg\phi)^M$.
11. $(\neg\Diamond\phi)^M = \Box(\neg\phi)^M$.

The definition may seem quite awkward, but in fact it is extremely natural: we simply rewrite a give formula so that all negations occur only in front of the propositional variables using de Morgan laws. We also eliminate all occurring implication symbols.

If $\phi = \phi^M$, then we say that ϕ is already in de Morgan form.

Now, instead of giving a proper definition of what we mean by tableaux for the propositional modal logic we will present a procedure of creating tableaux in a somewhat informal way. We have written above that tableaux are used to prove some propositional modal formulae. Actually we use this method to *refute* them. As a by-product of doing so, we construct a model in which the formula in question fails to hold.

So suppose we are given a formula ϕ in de Morgan form we would like to refute. We draw a **box** around the formula (some more formulae will be written inside the same box, so it is good to actually draw the box gradually).

At each point of the construction we will have a tree labelled with formulae, some of which will be grouped in boxes. Now, at every step we may enlarge the tree in the following way:

1. Whenever you see both p and $\neg p$ for some propositional variable p on the same branch and in the same world, you may write $\bot$ at the bottom of that branch in the same box as the last element of the branch. We say that this branch is **closed**.

2. Whenever at some vertex you see a formula $\phi \wedge \psi$ write ϕ and ψ subsequently at the end of each branch in the same box in which $\phi \wedge \psi$ is placed (i.e. at each vertex v in the same box as the node labelled with $\phi \wedge \psi$ that points to no element in the same box you draw two new nodes v' and v'' with arrows from v to v' and from v' to v''. If additionally v pointed at some nodes $v_1, \ldots v_n$ in another boxes you erase these old arrows going from v and draw new arrows from v'' to $v_1, \ldots v_n$). Cross out the instantiation of the formula $\phi \wedge \psi$ you used.

3. Whenever you see a node labelled with a formula $\phi \vee \psi$ you may draw at the same world at the end of each branch passing through the node with this occurrence of $\phi \vee \psi$ two new vertices v', v'' labelled with ϕ and ψ respectively and draw new arrows going from the last node v of the respective branch to v' and v'' (so that v has now two new children). If there were some arrows going from v to nodes $v_1, \ldots, v_n$ in other boxes, you erase these arrows draw a duplicate of each labelled tree starting at v_i and you draw arrows from v' to $v_1, \ldots, v_n$ and from v'' to the respective vertices in the duplicate (in practice you may simply draw arrows both from v' and v'' to $v_1, \ldots, v_n$). Cross out the instantiation of the formula $\phi \vee \psi$ you used.

4. Whenever you see a vertex v labelled $\Box\phi$, then for any node v' which lies in the same branch, but in another box joined by an arrow with the box in which the vertex v was located and which points at no vertex within the same box, you may draw a node v'' pointed at by v' and labelled with ϕ. If additionally v' was pointing at some vertices $v_1, \ldots, v_n$ in other boxes, then you may erase these arrows and draw the new arrows from v'' to $v_1, \ldots, v_n$.

5. Whenever you see a vertex v labelled with $\Diamond\phi$, then for any vertex v' in the same branch which points to no other vertex in the same branch you may draw a new box and a new vertex v'' pointed at by an arrow going from v' and labelled with ϕ. This is the only way in which new boxes are introduced.

You call a formula ϕ **Tableaux-refutable** iff there exists a tree formed according to the above rules, whose root is labelled with $(\neg\phi)^M$ and in which there is a branch which does not close. This means that there exists a Kripke structure in which the formula does not hold.

Now the following fact holds:

Fact 27. *A formula is provable iff it is not Tableaux-refutable.*

Exercise 9. Find Tableaux for the following formulae:

1. $\Box(p \vee q) \to (\Box p \vee \Box q)$
2. $p \to \Box\Diamond p$
3. $\Box(\Box p \to p) \to \Box p$
4. $\Box\Diamond p \to \Diamond\Box p$
5. $\Box(p \to \Diamond(p \wedge \Box p)) \to (\Box p \to \Diamond p)$
6. $\Box\Diamond(p \wedge q)$
7. $\Diamond(p \wedge \neg p) \to \Box q$
8. $\Diamond\Diamond p \to \Diamond p$
9. $\big(\Box\Box p \wedge \Diamond p\big) \to \Box p$
10. $\Diamond(p \wedge \Box(p \to q)) \to (\Box p \to \Box\Box q)$

2 First-Order Modal Logic

2.1 First-Order Logic

Syntax A (relational) *signature* (with constants) σ is a pair of nonempty sets $\mathcal{R}, \mathcal{C}$ called the sets of *relations* and *constants* respectively and a function $\rho : \mathcal{R} \to \mathbb{N}$ which assigns arities to symbols of $\mathcal{R}$. Intuitively ρ says how many objects can stand in the relation from $\mathcal{R}$.

We define relational formulae of the first-order logic by induction on complexity of formulae. Namely, we assume that we have fixed some set of first-order variables V which we will denote x, y, z and occasionally also $x_1, x_2, \ldots, y_1, y_2, \ldots, z_1, z_2, \ldots$. Then we define relational formulae of the fist-order logic over the language $\mathcal{L}$ as the smallest set F satisfying the following conditions:

1. $R(v_1, \ldots, v_n) \in F$, where v_i are either variables from V or constants from $\mathcal{C}$ and R is some relation symbol such that $\rho(R) = n$
2. $v_1 = v_2$ where v_1, v_2 are either variables from V or constants from $\mathcal{C}$.
3. $\phi \wedge \psi \in F$, when $\phi \in F$ and $\psi \in F$.
4. $\phi \vee \psi \in F$, when $\phi \in F$ and $\psi \in F$.

5. $\neg\phi \in F$, when $\phi \in F$.

6. $\forall v \ \ \phi$, when $\phi \in F, v \in V$.

7. $\exists v \ \ \phi$, when $\phi \in F, v \in V$.

Example 28. Suppose $\mathcal{R} = \{P, R\}$, $\mathcal{C} = \{c\}$, $\rho(P) = 1$, $\rho(R) = 2$. Then

$$R(x_1, c)$$

and

$$P(x_2)$$

are well built formulae. If so, then so are

$$R(x_1, c) \wedge P(x_2)$$

and

$$\exists x_1 \forall x_2 \big((R(x_1, c) \wedge P(x_2)) \big).$$

Semantics Let us recall how semantics for first-order logic is defined. Then we will try to use these ideas to define semantics for some first-order version of the modal logic.

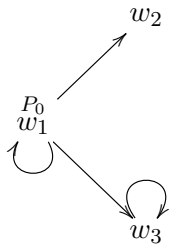
Definition 29. Let $\sigma = (\mathcal{R}, \mathcal{C}, \rho)$ be a signature. By a relational model $\mathcal{M}$ we mean any tuple (M, τ), where M is a nonempty set called the **domain** or the **universe** of M and τ is a function with domain $\mathcal{R} \cup \mathcal{C}$ such that

1. for every $R \in \mathcal{R}$, if $\rho(R) = n$, then $\tau(R) \subseteq M^n$

2. for every $c \in \mathcal{C}$, $\tau(c) \in M$.

A model can be thought of as a description of how many basic individuals exist (which objects form the universe) and how are the basic relations from our signature to be interpreted.

Example 30. Let $\sigma = (\mathcal{R}, \mathcal{C}, \rho)$ be such that $\mathcal{R} = \{R\} \cup \{P_i \mid i \in \mathbb{N}\}$, $\rho(R) = 2$, $\rho(P_i) = 1$ for every i and $\mathcal{C} = \emptyset$. Let $M = \{w_1, w_2, w_3\}$ and $\tau(R) = \{(w_1, w_2), (w_1, w_1), (w_3, w_3), (w_1, w_3)\}$ and $\tau(P_0) = a$, $\tau(P_i) = \emptyset$ for $i > 0$. Then M can be depicted as a Kripke model



with P_0 thought of as a propositional variable.

Let us define something more complex: working over the same signature put:

$$\begin{aligned} M &= \mathbb{N} \\ \tau(R) &= \{(m,n) \mid m < n\} \\ \tau(P_i) &= \{n \in \mathbb{N} \mid i \text{ divides } n\} \end{aligned}$$

Try drawing the above model.

When reading the following definition it is good to keep in mind that what we try to capture is our informal notion of a sentence true of some given structure. This is to be understood in a very straightforward manner in which $\forall x \exists y \;\; x < y$ is true in the set of natural numbers with its natural order but $\forall x \exists y \;\; y < x$ is not (why?).

Definition 31. By a valuation α on a first-order model (M, τ) we mean any function whose domain is the set of all first-order variables V and which takes values in the domain of M. It will be convenient to assume that α is defined also on constants from our signature and that for every $c \in \mathcal{C}$ we have $\alpha(c) = \tau(c)$. We define what does it mean for a formula ϕ to hold in a model M under valuation α, in symbols $M, \alpha \models \phi$, by induction on the complexity of formulae:

1. $M, \alpha \models R(v_1, \ldots, v_n)$ iff $(\alpha(v_1), \ldots, \alpha(v_n)) \in \tau(R)$ or, in other words, the elements assigned to $v_1, \ldots, v_n$ by α satisfy the relation R as interpreted in the model M.
2. $M, \alpha \models v_1 = v_2$ iff $\alpha(v_1) = \alpha(v_2)$.
3. $M, \alpha \models (\phi \wedge \psi)$ iff $M, \alpha \models \phi$ and $M, \alpha \models \psi$.
4. $M, \alpha \models (\phi \vee \psi)$ iff $M, \alpha \models \phi$ or $M, \alpha \models \psi$.

5. $M, \alpha \models \neg\phi$ iff it is not the case that $M, \alpha \models \phi$.

6. $M, \alpha \models \exists v\ \phi$ iff there exists a valuation α' which may differ from α at most in what it ascribes to the variable v such that $M, \alpha' \models \phi$.

7. $M, \alpha \models \forall v\ \phi$ iff $M, \alpha' \models \phi$ for all valuations α' which may differ from α at most in what it ascribes to the variable v.

The above inductive definition may seem scary, but it really captures the most intuitive notion of a sentence being true in some structures. We consider the more general case: namely formulae like $R(x, y) \wedge \exists z\ P(z)$. In order to say, whether this formula is satisfied or not in a given model, we have to specify what do we mean by x and y — and this is precisely what the valuations do: they are specification of what variables "mean". It might be confusing that we require valuations to be defined for *all* variables, not only the ones that actually occur in the formula, but this is a purely technical detail to simplify our considerations.

Additionally, we say that a formulae is **true** in a given model iff it is satisfied under all valuations. We denote it $M \models \phi$ (i.e. similarly to the above definition, but omitting the reference to a valuation). It is **valid** iff it is true in all models. Observe that if $\phi(x)$ is a formula in which x is not within the scope of any quantifier (we say that x is a *free variable* in ϕ) then it is true in a model M if and only if $\forall x\phi(x)$ is true in a model M.

Sometimes we will use the following convention: if ϕ contains exactly one free variable, then we instead of writing a model and a valuation on the lefthandside of the satisfiability relation we will be writing simply a model and an *individual* from the universe, meaning that ϕ is satisfied in a model by the valuation which assigns the chosen individual to the only free variable of ϕ.

Example 32. Let (M, τ) be the first model from Example 30. Then $(M, \tau) \models \exists x(R(x, x) \wedge P_0(x))$. Indeed, let α be an arbitrary valuation and α' be such that

$$\begin{aligned} \alpha'(x) &= w_1 \\ \alpha'(y) &= \alpha(y) \text{ for } y \neq x \end{aligned}$$

Then $(M, \tau), \alpha' \models R(x, x) \wedge P_0(x)$. Indeed, the latter holds if and only if $(\alpha'(x), \alpha'(x)) \in \tau(R)$ and $\alpha'(x) \in \tau(P_0)$ which is true since $\alpha'(x) = w_1$.

Let now (M, τ) be the second model from the same example. We shall show that $(M, \tau) \models \forall x \exists y\big(R(x, y) \wedge P_3(y)\big)$. Let α be an arbitrary valuation.

Let α' be an valuation which differs from α at most on the value assigned to x. We shall show

$$(M, \tau), \alpha' \models \exists y\big(R(x, y) \wedge P_3(y)\big)$$

So we have to find a way of assigning the value to variable y such that for the resulting valuation α'' we have

$$(M, \tau), \alpha'' \models \big(R(x, y) \wedge P_3(y)\big)$$

The above is true if and only (by the definition of our model!) $\alpha''(x) < \alpha''(y)$ and $3|\alpha''(y)$. So for example we can put $\alpha''(y) = 3(\alpha(x) + 1)$.

It is easy to see that the sentence $\exists x \;\; R(x, x)$ is not true in the above model. However this sentence is true in the first model from Example 30, since for example $(w_1, w_1) \in \tau(R)$.

Exercise 10. Let $\sigma = (\mathcal{R}, \mathcal{C}, \rho)$, where $\mathcal{R} = \{R\}$ and $\rho(R) = 2$. Check whether given sentences are true in given models

1. Let $M = \mathbb{N}$ and $(m, n) \in \tau(R)$ if and only if $m < n$.
   (a) $\exists x \forall y R(x, y)$
   (b) $\forall y \exists x R(y, x)$
   (c) $\forall y \exists x R(x, y)$
2. $M = \mathbb{N}$ and $(m, n) \in \tau(R)$ if and only if $m > n$.
   (a) $\exists x \forall y R(x, y)$
   (b) $\forall y \exists x R(y, x)$
   (c) $\forall y \exists x R(x, y)$
3. $M = \mathbb{N}$ and $(m, n) \in \tau(R)$ if and only if m divides n.
   (a) $\exists x \forall y R(x, y)$
   (b) $\exists x \exists y\big(\neg(x = y) \wedge \forall z(\neg R(z, x) \wedge \neg R(z, y)\big)$
4. Let (M, τ) be the first model from Example 30.
   (a) $\forall x \exists y(R(x, y) \wedge P_0(y))$
   (b) $\exists x \exists y(\neg(x = y) \wedge R(x, x) \wedge R(y, y))$
   (c) $\exists x \forall y R(x, y)$.

2.1.1 The standard translation

We will show that the propositional modal logic (pml) is a part of the classical first order logic (cfol). Obviously this is not literally true, since both logic use different symbols. However we can define a translation * of formulae of pml to formulae of cfol. To do this it will be convenient to assume that the only propositional variables we use in formula of modal logic are of the form p_i, where i is a natural number. This translation will be truth preserving in the following sense: if ϕ is a formula of pml, then its translation, denoted ϕ^* will be a formula of cfol over a signature

$$(\{R'\}, \{P_i \mid i \text{ is a natural number }\}), \rho(R) = 2, \rho(P_i) = 1$$

(R will represent the accessibility relation, while $P_i's$- propositional variables) with precisely one free variable x such that for arbitrary Kripke model $\mathcal{K} = (K, R, f)$ and a world $w \in K$ we will have

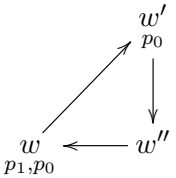
$$\mathcal{K}, w \models \phi \text{ if and only if } \mathcal{K}^*, w \models \phi^*.$$

Where $\mathcal{K}^*$ denotes the natural counterpart of $\mathcal{K}$ in cfol, i.e. the model (K, τ) such that

$$\begin{aligned} \tau(R') &= R \\ \tau(P_i) &= \{w \in K \mid \mathcal{K}, w \models p_i\} \end{aligned}$$

Let us unravel this definition: $\mathcal{K}^*$ results from $\mathcal{K}$ by taking possible worlds (i.e. elements of K) to be individuals (i.e. the universe of Kripke model form now the universe of first order model) and interpreting R' as the accessibility relation. Moreover each P_i is interpreted as the set of those worlds w which are labelled with p_i in $\mathcal{K}$.

Example 33. Let $\mathcal{K} = (K, R, f)$ be the following model



then $\mathcal{K}^* = (K, \tau)$ and

$$\begin{aligned} \tau(R') &= \{(w, w'), (w', w''), (w'', w)\}, \\ \tau(P_0) &= \{w, w'\}, \\ \tau(P_1) &= \{w\}, \\ \tau(P_i) &= \emptyset, \text{ for } i > 1. \end{aligned}$$

To define the translation it will be convenient to assume that the individual variables we use are numbered with natural numbers: we will denote them by $x_0, x_1, x_2 \ldots$. Let us finally define *. The translation will be *compositional*, i.e. the translation of a formula will be fully determined by its main connective (or modal operator) and the translations of its immediate subformulae. We will define it by induction on the structure of a formula: we start with the simplest formulae possible, i.e. propositional variables, and then show how to compute the translations of more complex formulae once the translations of their immediate subformulae have been fixed.

Definition 34 (The standard translation)**.** For every p_i,

$$(p_i)^* = P_i(x_0).$$

If the translation of ψ has already been defined, then

$$(\neg\psi)^* = \neg(\psi)^*.$$

If the translations of ψ, θ have already been defined, then

$$\begin{aligned}(\theta \wedge \psi)^* &= (\theta)^* \wedge (\psi)^* \\ (\theta \vee \psi)^* &= (\theta)^* \vee (\psi)^* \\ (\theta \to \psi)^* &= (\theta)^* \to (\psi)^*\end{aligned}$$

Suppose now the translation of ψ has been defined and let n be the number of modal operators used in ψ. Then

$$\begin{aligned}(\Diamond\psi)^* &= \exists x_{n+1}\big(R(x_0, x_{n+1}) \wedge (\psi)^*[x_{n+1}/x_0]\big) \\ (\Box\psi)^* &= \forall x_{n+1}\big(R(x_0, x_{n+1}) \to (\psi)^*[x_{n+1}/x_0]\big)\end{aligned}$$

where $\phi[x_{n+1}/x_0]$ denotes the result of replacing in ϕ every occurrence of variable x_0 with variable x_{n+1}.

It would be best to see how this definition works on a concrete example:

Example 35. We will compute the translation of $\Box\Diamond p_0 \to \Diamond\Box p_0$. The main connective in this formula is the implication, so

$$(\Box\Diamond p_0 \to \Diamond\Box p_0)^* = (\Box\Diamond p_0)^* \to (\Diamond\Box p_0)^*$$

Let us compute $(\Box\Diamond p_0)^*$ first. The main connective is $\Box$ and $\Diamond p_0$ contains exactly one modal operator. So

$$(\Box\Diamond p_0)^* = \forall x_2\big(R(x_0, x_2) \to (\Diamond p_0)^*[x_2/x_0]\big).$$

So let us compute $(\Diamond p)^*$: p contains no modal operators, so

$$\begin{array}{rcl}(\Diamond p_0)^* & = & \exists x_1\big(R(x_0,x_1)\wedge(p_0)^*[x_1/x_0]\big)\\ & = & \exists x_1\big(R(x_0,x_1)\wedge(P_0(x_0))[x_1/x_0]\big)\\ & = & \exists x_1\big(R(x_0,x_1)\wedge P_0(x_1)\big)\end{array}$$

Now we have to plug in the last formula changing x_0 to x_2. We get:

$$(\Box\Diamond p_0)^* = \forall x_2\big(R(x_0,x_2)\to\exists x_1\big(R(x_2,x_1)\wedge P_0(x_1)\big)\big).$$

Similarly

$$(\Diamond\Box p_0)^* = \exists x_2\big(R(x_0,x_2)\wedge\forall x_1\big(R(x_2,x_1)\to P_0(x_1)\big)\big)$$

So the translation of $\Box\Diamond p_0\to\Diamond\Box p_0$ is the following formula:

$$\Big(\forall x_2\big(R(x_0,x_2)\to\exists x_1\big(R(x_2,x_1)\wedge P_0(x_1)\big)\big)\Big)\longrightarrow$$
$$\longrightarrow\Big(\exists x_2\big(R(x_0,x_2)\wedge\forall x_1\big(R(x_2,x_1)\to P_0(x_1)\big)\big)\Big)$$

Exercise 11. Compute translations of the following formulae:

1. $\Box p_3$
2. $\Diamond(p_1\wedge p_{17})$
3. $\Box(p_0\vee\neg\Diamond p_{13})$
4. $\Box\neg p_2\vee\Box p_2$
5. $\Box\Diamond p_0\to\Diamond\Box p_0$
6. $\Box(\Box p_0\to p_0)\to\Box p_0$
7. $\Box(p_3\wedge p_2)\to\big(\neg\Diamond\Box P_5\big)$.
8. $\Box(p_4\vee\neg\neg\Diamond p_3)$.

3 Quantified modal logic

3.1 Models for the quantified modal logic

We extend the definition of first-order formulae in a natural way, so that our formulae may contain some modal operators. Let σ be a relational signature and V an infinite set of *variables*.The set of relational *first-order modal formulae* is the smallest set F satisfying the following conditions:

1. $R_i(v_1, \ldots, v_n) \in F$, where v_i are variables and R_i is some relation symbol from σ of arity n.
2. $\phi \wedge \psi \in F$, when $\phi \in F$ and $\psi \in F$.
3. $\phi \vee \psi \in F$, when $\phi \in F$ and $\psi \in F$.
4. $\neg\phi \in F$, when $\phi \in F$.
5. $\forall v \;\; \phi \in F$, when $\phi \in F, v \in V$.
6. $\exists v \;\; \phi \in F$, when $\phi \in F, v \in V$.
7. $\Box\phi \in F$, when $\phi \in F$.
8. $\Diamond\phi \in F$, when $\phi \in F$.

As before, unwinding this definition, we see that quantified modal formulae look like the regular first-order formulae, but we allow some extra operators $\Box$ and $\Diamond$.

Now we are ready to introduce the models for the first order modal logic.

Definition 36. By a Kripke model $\mathcal{W}$ for first-order modal logic over the signature σ we mean the structure (W, R, D, V) such that:

1. W is a nonempty set whose elements we call **possible worlds** like in the propositional modal case.
2. R is a binary relation on W, which we call the **accessibility relation** as in the propositional case.
3. D is a function with domain W such that for every $w \in W$, $D(w)$ is a nonempty set. We think of $D(w)$ as of the domain of the possible world associated with w. The set theoretical sum of the set of values of D will be called **the set of individuals** of $\mathcal{W}$. I.e. the set of individuals of $\mathcal{W}$ is
$$\bigcup_{w \in W} D(w)$$

4. V is a function which takes a world w and a symbol h from the signature and returns the interpretation of h in $D(w)$.

Remark 37. If (W, R, D, V) is a Kripke model for FOML, then for every $w \in W$ $(D(w), V(w, \cdot))$ is a model for first order logic, as in Definition 29. ($V(w, \cdot)$ denotes the function resulting from V by fixing one of its arguments)

Again, we intuitively think of the elements $w \in W$ as of the possible worlds. As before, the relation R holds between w and v if, intuitively, world v is possible *from the point of view* of world w.

Definition 38. Let $\mathcal{W} = (W, R, D, V)$ be a FOML model. A **valuation** α is any function from the chosen set of first order variables to the set of individuals of $\mathcal{W}$. For every $w \in W$ and every FOML formula ϕ we call α a w-**valuation with respect to** ϕ if for every free variable v in ϕ, $\alpha(v)$ belongs to $D(w)$.

Now, we define what is a satisfaction of a formula in a given model for FOML.

Definition 39. Let $\mathcal{W} = (W, R, D, V)$ be any Kripke model over a signature σ. We define what does it mean for a formula ϕ to be satisfied in a world $w \in W$ in the model $\mathcal{W}$ under a valuation α by induction on complexity of formulae.

1. If $\phi = R(x_1, \ldots, x_n)$, then $\mathcal{W}, w, \alpha \models \phi$ iff $(\alpha(x_1), \ldots, \alpha(x_n)) \in V(w, R)$
2. $\mathcal{W}, w, \alpha \models (\phi \wedge \psi)$ iff $\mathcal{W}, w, \alpha \models \phi$ and $\mathcal{W}, w, \alpha \models \psi$.
3. $\mathcal{W}, w, \alpha \models (\phi \vee \psi)$ iff $\mathcal{W}, w, \alpha \models \phi$ or $\mathcal{W}, w, \alpha \models \psi$.
4. $\mathcal{W}, w, \alpha \models \neg\phi$ iff it is not the case that $\mathcal{W}, w, \alpha \models \phi$.
5. $\mathcal{W}, w, \alpha \models \exists v\, \phi$ iff there exists a w-valuation with respect to ϕ α' which may differ from α at most in what it ascribes to variable v such that $\mathcal{W}, w, \alpha' \models \phi$.
6. $\mathcal{W}, w, \alpha \models \forall v\, \phi$ iff $\mathcal{W}, w, \alpha' \models \phi$ for all w-valuations with respect to ϕ α' which differ from α at most in what they ascribe to variable v.
7. $\mathcal{W}, w, \alpha \models \Box\phi$ iff $\mathcal{W}, v, \alpha \models \phi$ for all worlds v such that $R(w, w')$ and α is a w' valuation with respect to ϕ.
8. $\mathcal{W}, w, \alpha \models \Diamond\phi$ iff $\mathcal{W}, v, \alpha \models \phi$ for some world v such that $R(w, w')$ and α is a v valuation with respect to ϕ.

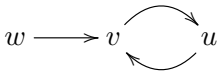
Similarly to the first-order case, we say that a formula over a signature σ is true in a Kripke model, at a given world w iff it is satisfied there under all valuations. We say that it is true in a model $\mathcal{W}$, if it is true at all worlds in $\mathcal{W}$, we say that it is valid, if it is true in all models over the signature σ at all worlds.

Exercise 12. Check whether the following formulae hold in every Kripke model.

1. $\big(\exists x \Diamond P(x)\big) \to \exists x P(x)$.
2. $\big(\Diamond \exists x P(x)\big) \to \big(\exists x \Diamond P(x)\big)$.
3. $\big(\forall x \Box P(x)\big) \to \big(\Box \forall x P(x)\big)$.
4. $\big(\Diamond \forall x P(x)\big) \to \big(\forall x \Diamond P(x)\big)$.
5. $\big(\Diamond \forall x \Box P(x)\big) \to \big(\Diamond \Box \forall x P(x)\big)$.
6. $\big(\exists x \Box P(x)\big) \to \big(\Box \exists x P(x)\big)$.
7. $\big(\Diamond \forall x \exists y (Q(x,y))\big) \to \big(\forall x \Diamond \exists y Q(x,y)\big)$
8. $\big(\Box \exists x \forall y Q(x,y)\big) \to \big(\exists x \Box \forall y Q(x,y)\big)$
9. $\Big(\big(\exists x \Diamond P(x)\big) \wedge \big(\Box \forall x (P(x) \to F(x))\big)\Big) \to \big(\exists x \Diamond F(x)\big)$
10. $\big(\Box \forall x (P(x) \to F(x)) \wedge \Box \forall x (F(x) \to G(x))\big) \to \forall x (\Box P(x) \to \Box G(x))$
11. $\exists x \exists y (Q(x,y) \wedge \Box Q(y,x))$

Exercise 13. Check whether the following formulae hold in the given models:

1. Let W and R be as presented:



   (a) Define the Domain and the interpretation functions as

$$D(w) = \{a,b,c\}, D(v) = \{a,b\}, D(u) = \{b,c\}$$

$$V(P,w) = \{a,b,c\}, V(P,v) = \{a,b\}, V(P,u) = \{b\}.$$

   Verify whether the following formulae are satisfied in $(W,R,D,V),w$

i. $\Box\forall x P(x)$
ii. $\forall x \Box P(x)$
iii. $\forall x \Box\Box P(x)$
iv. $\exists x(P(x) \wedge \Diamond\neg P(x))$
v. $\exists x(P(x) \wedge \Diamond\Diamond\neg P(x))$

(b) Define the domain and the interpretation functions as

$$D(w) = \{a, b\}, D(v) = \{a, b, c\}, D(u) = \{b, c\}$$

$$V(P, w) = \{a, b\}, V(P, v) = \{a, b\}, V(P, u) = \{b\}.$$

Verify whether the following formulae are satisfied in $(W, R, D, V), w$

i. $\Diamond\exists x\Diamond\neg P(x)$
ii. $\forall x \Box P(x) \rightarrow \Box\forall x \Box P(x)$

(c) Define the Domain and the interpretation functions as

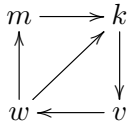
$$D(w) = \{a\}, D(v) = \{a, b\}, D(u) = \{a, b, c\}$$

$$V(P, w) = \{a, b, c\}, V(P, v) = \emptyset, V(P, u) = \{b\}.$$

Verify whether the following formulae are satisfied in $(W, R, D, V), u$ (mind that we switched the world from which we start)

i. $\exists x \Box P(x) \rightarrow \Box\exists x P(x)$.

2. Let the domain D and the relation R be now as follows:



Define the interpretation function D and the interpretation function V as follows:

(a)

$$D(w) = \{a, b\}, D(m) = \{a\}, D(k) = \{b\}, D(v) = \{a, b, c\}$$

$$V(Q, w) = \{(a, b), (b, a)\}, V(Q, m) = \{(a, a)\},$$
$$V(Q, k) = \{(b, b)\}, V(Q, v) = \{(a, b), (a, c), (b, c)\}$$

Verify whether the following formulae holds in $(W, R, D, V), w$

i. $\forall x \Box Q(x,x) \to \forall x Q(x,x)$

ii. $\forall x \Box Q(x,x) \to \Box \Diamond \forall x \exists y Q(x,y)$

(b)

$$D(w) = \{a,b\}, D(m) = \{a\}, D(k) = \{b,c\}, D(v) = \{a,b,c\}$$

$$V(Q,w) = \{(a,b),(b,a)\}, V(Q,m) = \{(a,a)\},$$
$$V(Q,k) = \{(b,c)\}, V(Q,v) = \{(a,b),(a,c),(b,c)\}$$

Verify whether the following formulae holds in $(W,R,D,V),w$

i. $\exists x \Box Q(x,x)$

ii. $\forall x \Box Q(x,x)$

iii. $\big(\forall x \forall y (Q(x,y) \equiv Q(y,x))\big) \to \Diamond \big(\forall x \forall y (Q(x,y) \equiv Q(y,x))\big)$

iv. $\big(\forall x \forall y (Q(x,y) \equiv Q(y,x))\big) \to \Box \big(\forall x \forall y (Q(x,y) \equiv Q(y,x))\big)$

Exercise 14. Solve the following puzzle of Quine:

> Cyclists are necessarily two-legged, but not necessarily rational. Mathematicians are necessarily rational, but not necessarily two-legged. Consider a cycling mathematician. Is he both necessarily rational and not necessarily rational, with the same contradiction in his legs?

Quotation after "Modal Logic for Open Minds".

4 Intuitionistic logic

Intuitionistic logic aims to capture patterns of constructive reasoning. In a way, we assume that nothing is either true or false, until proven to be so. In particular we do not assume that the law of excluded middle holds.

4.1 Propositional intuitionistic logic

The syntax of propositional intuitionistic logic is almost the same as the syntax of the classical propositional logic. We take $\vee$, $\wedge$, $\to$ and $\bot$ as primitive symbols, where this last one is a propositional constant. For example

$$(p \to \bot) \wedge (q \wedge (r \vee \bot))$$

is a well-built formula. The semantics for Intuitionistic Logic is completely different. We will translate intuitionistic formulae into some specific class of modal formulae and then define satisfaction of the intuitionistic formulae in Kripke models via this translation.

Definition 40. We define a translation of a propositional intuitionistic formula ϕ to a modal formula ϕ^T by induction on complexity in the following way:

1. $\phi^T = \phi$ for ϕ propositional variables.
2. $(\phi \wedge \psi)^T = \phi^T \wedge \psi^T$.
3. $(\phi \vee \psi)^T = \phi^T \vee \psi^T$.
4. $(\phi \to \psi)^T = \Box(\phi^T \to \psi^T)$.
5. $(\bot)^T = (p \wedge \neg p)$

Definition 41 (Negation)**.** We treat $\neg$ as defined symbol. We put

$$\neg\phi := \phi \to \bot$$

Note that the formula $\neg\phi$ translates into a modal formula $(\neg\phi)^T$ as $\Box(\phi^T \to (p \wedge \neg p))$ for some propositional variable p. In other words, $\neg\phi$ is interpreted in intuitionistic logic as *necessarily* not ϕ.

Now we define the class of models for the intuitionistic propositional logic. It will consist of Kripke models of particular kind.

Definition 42. Let $\mathcal{K} = (K, R, V)$ be a Kripke model for the propositional modal logic. We call M a model for the propositional intuitionistic logic iff the following additional conditions are satisfied:

1. R is a partial order, i.e. it is antisymmetric, reflexive and transitive.
2. V is monotonous, i.e. for every w, v such that wRv, if $\mathcal{K}, w \models p$, then $\mathcal{K}, v \models p$ for every propositional variable p.

Now we define, what does it mean for an intuitionistic propositional formula to be defined in a Kripke model.

Definition 43. Let ϕ be an intuitionistic propositional formula, let $\mathcal{K}$ be a Kripke model for the propositional intuitionistic logic and let w be a world in K. We say that ϕ is *satisfied* in the model $\mathcal{K}$ at the world w, denoted $\mathcal{K}, w \models_i \phi$ iff

$$\mathcal{K}, w \models \phi^T$$

where $\models$ is a satisfaction relation for Propositional Modal Logic.

Remark 44 (Satisfaction conditions for Propositional Intuitionistic Logic)**.** Let us note that $\models_i$ satisfies:

1. $\mathcal{K}, w \models_i p$ iff $\mathcal{K}, w \models p$
2. $\mathcal{K}, w \models_i \phi \vee \psi$ iff $\mathcal{K}, w \models_i \phi$ or $\mathcal{K}, w \models_i \psi$
3. $\mathcal{K}, w \models_i \phi \wedge \psi$ iff $\mathcal{K}, w \models_i \phi$ and $\mathcal{K}, w \models_i \psi$
4. $\mathcal{K}, w \models_i \phi \to \psi$ iff for every v such that wRv, if $\mathcal{K}, v \models_i \phi$, then $\mathcal{K}, v \models_i \psi$
5. it is never true that $\mathcal{K}, w \models_i \bot$

From the last two it holds that

$$\mathcal{K}, w \models_i \neg\phi \text{ iff for every } v \text{ such that } wRv, \mathcal{K}, v \not\models_i \phi$$

Definition 45. We say that a formula ϕ *holds* in a model $\mathcal{K} = \langle K, R, V\rangle$, $\mathcal{K} \models_i \phi$, if for every $w \in K$ it holds that $\mathcal{K}, w \models_i \phi$. We say that a formula ϕ is *intuitionistically valid* if for every $\mathcal{K}$ ϕ holds in $\mathcal{K}$. We use $\models_i \phi$ to denote that ϕ is valid.

Convention 46. When drawing models, to avoid drawing many arrows, we implicitly assume that the relation is reflexive and transitive. Hence every time we have

$$w \longrightarrow v \longrightarrow x$$

we mean that also the following pairs are in the accessibility relation:

1. $\langle w, w\rangle$
2. $\langle v, v\rangle$
3. $\langle x, x\rangle$
4. $\langle w, x\rangle$

Exercise 15. Verify whether given models satisfy given formulae of intuitionistic propositional logic:

1. Model (p, q denote propositional variables and w, v, x denote worlds)

$$\underset{p}{w} \longrightarrow \underset{p,q}{v} \longrightarrow \underset{p,q}{x}$$

   Check whether

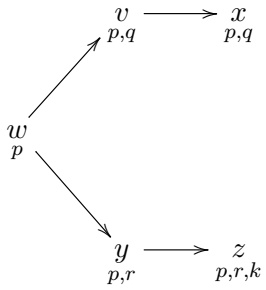
(a) $\mathcal{K}, w \models_i p \to q$

(b) $\mathcal{K}, w \models_i \neg p$

(c) $\mathcal{K}, w \models_i \neg q$

(d) $\mathcal{K}, w \models_i \neg\neg q$

2. Model: (p, q, r, k denote propositional variables and w, v, x, y, z denote worlds)



Formulae:

(a) $\mathcal{K}, w \models_i r \vee \neg r$

(b) $\mathcal{K}, w \models_i q \to p$

(c) $\mathcal{K}, w \models_i \neg q$

(d) $\mathcal{K}, y \models_i \neg q$

(e) $\mathcal{K}, v \models_i p \to q$

(f) $\mathcal{K}, w \models_i (p \to q) \to q$

Exercise 16. Check whether the following formulae of propositional intuitionistic logic are intuitionistically valid:

1. $p \to p$
2. $(p \to q) \to p$
3. $p \to (\neg\neg p)$
4. $(\neg\neg p) \to p$
5. $(p \wedge q) \to p$
6. $(p \wedge \neg p) \to q$

4.1.1 Hilbert-style proof system for IPL and the Disjunction Property

Let us define the Hilbert-style proof system for Intuitionistic Propositional Logic. The notion of *proof* and *provability* is the same as in the classical case, see Definition 5 and we take Modus Ponens as our unique rule of reasoning. In contrast to classical case we take the following axiom schemes (i.e. every formula of one of the following shapes is our axiom):

1. $\phi \to (\psi \to \phi)$
2. $(\phi \to (\psi \to \chi)) \to ((\phi \to \psi) \to (\phi \to \chi))$
3. $\phi \to \phi \vee \psi$
4. $\psi \to \phi \vee \psi$
5. $(\phi \to \chi) \to ((\psi \to \chi) \to (\phi \vee \psi \to \chi))$
6. $\phi \wedge \psi \to \psi$
7. $\phi \wedge \psi \to \phi$
8. $\phi \to (\psi \to \phi \wedge \psi)$
9. $\bot \to \phi$

The above system is taken from Proof Theory lecture notes by Benno van den Berg (available here). Let us use $\vdash_i \phi$ to denote the fact that ϕ is provable in the above defined system. Then we have

Theorem 47 (Completeness for IPL)**.** *For every formula ϕ, $\vdash_i \phi$ if and only if $\models_i \phi$.*

On of the distinctive virtues of intuitionistic logic, which witness its constructive character, is the following Disjunction Property:

Theorem 48 (Disjunction Property)**.** *For every formulae ϕ, ψ we have*

$$\vdash_i \phi \vee \psi \text{ if and only if } \vdash_i \phi \text{ or } \vdash_i \psi$$

We shall prove the above using the following monotonicity condition for intuitionistic Kripke models: it generalizes condition 2 in Definition 42 to arbitrary formulae:

Lemma 49. *Let ϕ be an intuitionistic formula, $\mathcal{K} = \langle K, R, V \rangle$ an intuitionistic Kripke model. Then we have: for all $w, v \in K$ such that wRv*

$$\mathcal{K}, w \models \phi \text{ implies } \mathcal{K}, v \models \phi \tag{MON}$$

Proof. We use induction on the complexity of formulae.

Base step If p is a propositional variable, then for every $w, v \in K$ such that wRv, MON holds by the definition of Intuitionistic Kripke Model.

Induction step Suppose ϕ is a compound formula and let us distinguish cases.

Case 1 Suppose $\phi = \psi_0 \vee \psi_1$ and MON holds for ψ_0 and ψ_1 and arbitrary $w, v \in K$ such that wRv. Let us fix $w, v \in K$ such that wRv and suppose

$$\mathcal{K}, w \models \psi_0 \vee \psi_1$$

then by definition either $\mathcal{K}, w \models \psi_0$ or $\mathcal{K}, w \models \psi_1$. Without loss of generality suppose the former holds (if the latter holds, then the proof is the same). By induction assumption for ψ_0 we get

$$\mathcal{K}, v \models \psi_0$$

Hence $\mathcal{K}, v \models \psi_0 \vee \psi_1$ and this step is finished.

Case 2 Suppose $\phi = \psi_0 \wedge \psi_1$. The proof is as above and we leave it as an exercise.

Case 3 Suppose $\phi = \psi_0 \to \psi_1$ and MON holds for ψ_0, ψ_1 and arbitrary w, v such that wRv. Let us fix $w, v \in K$ such that wRv. Assume that

$$\mathcal{K}, w \models \psi_0 \to \psi_1.$$

We have to check whether $\mathcal{K}, v \models \psi_0 \to \psi_1$. This amounts to checking whether for arbitrary u such that vRu

$$\text{if } \mathcal{K}, u \models \psi_0 \text{ then } \mathcal{K}, u \models \psi_1. \qquad (*)$$

So let us fix arbitrary u such that vRu. We have wRv and vRu, hence, by transitivity wRu. Hence $*$ follows by our assumption that $\mathcal{K}, w \models \psi_0 \to \psi_1$. □

Now we may proceed to the proof of Theorem 48:

Proof of Theorem 48. Let us assume that $\nvdash_i \phi$ and $\nvdash_i \psi$. By Theorem 47 there exists $\mathcal{K} = \langle K, R, V \rangle$, $w \in K$ such that

$$\mathcal{K}, w \not\models \phi$$

and $\mathcal{K}' = \langle K', R', V \rangle$ and $w' \in K'$ such that

$$\mathcal{K}', w' \not\models \psi.$$

Without loss of generality assume that $K \cap K' = \emptyset$ and take any $w'' \notin K \cup K'$. Let us define $\mathcal{K}'' = \langle K'', R'', V'' \rangle$ by

$$\begin{aligned} K'' &= \{w''\} \cup K \cup K' \\ R'' &= \{\langle w'', w \rangle, \langle w'', w' \rangle\} \cup R \cup R' \\ V'' &= V \cup V' \end{aligned}$$

($\mathcal{K}''$ results from "glueing" together models $\mathcal{K}$ and $\mathcal{K}'$ using the world w''). Let us observe that for every formula θ we have

$$\mathcal{K}'', w \models \theta \iff \mathcal{K}, w \models \theta \qquad (*)$$

and

$$\mathcal{K}'', w' \models \theta \iff \mathcal{K}', w' \models \theta. \qquad (**)$$

We claim that $\mathcal{K}'', w'' \not\models \phi \vee \psi$, which would end our proof. Indeed, for suppose the contrary. Without loss of generality assume that $\mathcal{K}'', w'' \models \phi$. Then by Lemma 49 we have that $\mathcal{K}'', w \models \phi$, and by $*$

$$\mathcal{K}, w \models \phi$$

which contradicts our assumption. □

4.2 Intuitionistic First-Order Logic

The syntax of Intuitionistic First-Order Logic is the same as the syntax of Classical First Order Logic, except for we take all of $\wedge$, $\vee$, $\rightarrow$, $\bot$, $\forall$ and $\exists$ as primitive symbols. Let us define the semantics for this logic: we start with the definition of a submodel.

Definition 50 (Submodel)**.** Let $\mathcal{M} = (M, \tau_M)$ and $\mathcal{N} = (N, \tau_N)$ be two first order models over constant-relational signature σ in the sense of Definition 29. We say that $\mathcal{M}$ is a *submodel* of $\mathcal{N}$ if and only if

1. for every relation $R \in \sigma$, $\tau_M(R) \subseteq \tau_N(R)$,
2. for every constant $c \in \sigma$, $\tau_M(c) = \tau_N(c)$.

If $\mathcal{M}$ is a submodel of $\mathcal{N}$ then we shall denote it by $\mathcal{M} \subseteq \mathcal{N}$.

Example 51. Let σ consists of a unary predicate P and a constant c. Let

$$\begin{aligned} M = \{x, y\}, \qquad & N = \{x, y, z\} \\ \tau_M(P) = \{x\}, \qquad & \tau_N(P) = \{x, y, z\} \\ \tau_M(c) = x, \qquad & \tau_N(c) = x \end{aligned}$$

Then $\mathcal{M} = (M, \tau_{\mathcal{M}})$ is a submodel of $\mathcal{N} = (N, \tau_{\mathcal{N}})$.

If we altered this definition putting $\tau'_M(P) = \{x, y\}$ and $\tau'_N(P) = \{x, z\}$, then (M, τ'_M) would not be a submodel of (N, τ'_N).

As in the case of propositional logic, we shall define models for Intuitionistic First-Order Logic as particular Kripke models for First-Order Modal Logic. It will be convenient to use one convention

Convention 52. In Definition 42 we defined models for First-Order Modal Logic as quadruples $\langle W, R, D, V\rangle$ such that D is the domain function and V is the interpretation function, such that for each $w \in W$,

$$\langle D(w), V(w, \cdot)\rangle$$

is a model for First-Order Logic ($V(w, \cdot)$ denotes the function of one argument resulting from V by fixing one particular world w). Equivalently we can say that we have a family $\{\mathcal{A}_w\}_{w\in W}$ of models for First-Order Logic parametrized by elements of W and define First Order Kripke models as

$$\langle W, R, \{\mathcal{A}_w\}_{w\in W}\rangle$$

where,as previously, W is a non-empty set, R is a binary relation ("accessibility" relation) and $\{\mathcal{A}_w\}_{w\in W}$ is a family of models for First-Order Logic parametrized by elements of W. We will use this definition.

Definition 53. An Kripke models for Intuitionistic First Order Logic is a triple $\langle W, R, \{A\}_{w\in W}\rangle$ where

1. W is a non-empty set.
2. $R \subseteq W^2$ is a partial order.
3. $\langle W, R, \{\mathcal{A}\}_{w\in W}\rangle$ is a family of models for First Order Logic over the same constant-relational signature σ such that for every $w, v \in W$ such that wRv we have
$$\mathcal{A}_w \subseteq \mathcal{A}_v$$

Remark 54. Let $\mathcal{W} = \langle W, R, \{\mathcal{A}_w\}_{w \in W} \rangle$ be a Kripke model for First-Order Intuitionistic Logic, $w \in W$ and let ϕ be a formula (of First-Order Logic). Recall the notion of w-valuation with respect to ϕ introduced in Definition 38. Observe that for every v such that wRv, if α is a w-valuation with respect to ϕ, then it is also v-valuation with respect to ϕ.

Convention 55. Let α be any valuation and x - a variable. For every b, by $\alpha[x \mapsto b]$ we denote the unique valuation β defined

$$\begin{aligned} \beta(y) &= \alpha(y) \text{ for } y \neq x \\ \beta(x) &= b \end{aligned}$$

I.e. $\alpha[x \mapsto b]$ differs from α at most on the value assigned to x, and $\alpha[x \mapsto b]$ assigns b to x.

Definition 56 (Satisfaction Relation)**.** Let $\mathcal{W} = \langle W, R, \{\mathcal{A}_w\}_{w \in W} \rangle$ be a Kripke model for First-Order Intuitionistic Logic. For every $v \in W$, let $\mathcal{A}_v = \langle A_v, \tau_v \rangle$. By induction on the complexity of first order formula ϕ we define the relation

$$\mathcal{W}, w, \alpha \models_i \phi$$

where $w \in W$ and α is a w valuation with respect to ϕ.

1. if $\phi = P(x_{i_0}, \ldots, x_{i_n})$, where x_{i_k} are either free variables or constants and P is an n-ary relational symbol from the signature, then

   $$\mathcal{W}, w, \alpha \models_i P(x_{i_0}, \ldots, x_{i_n})$$

   iff $\langle a_{i_0}, \ldots, a_{i_n} \rangle \in \tau_w(P)$ where for each $k \leq n$,

   $$a_{i_k} = \begin{cases} \tau_w(x_{i_k}), \text{ if } x_{i_k} \text{ is a constant,} \\ \alpha(x_{i_k}), \text{ if } x_{i_k} \text{ is a variable} \end{cases}$$

2. similarly if ϕ is of the form $x_{i_0} = x_{i_1}$ then

   $$\mathcal{W}, w, \alpha \models_i x_{i_0} = x_n$$

   iff $a_0 = a_1$ where a_l (for $l \leq 1$) equals $\tau_w(x_{i_l})$ iff x_{i_l} is a constant and $\alpha(x_{i_l})$ iff x_{i_l} is a variable.

3. if $\phi = \psi_0 \wedge \psi_1$, then

   $$\mathcal{W}, w, \alpha \models_i \psi_0 \wedge \psi_1$$

   iff $\mathcal{W}, w, \alpha \models_i \psi_0$ and $\mathcal{W}, w, \alpha \models_i \psi_1$.

4. if $\phi = \psi_0 \vee \psi_1$, then
$$\mathcal{W}, w, \alpha \models_i \psi_0 \vee \psi_1$$
iff $\mathcal{W}, w, \alpha \models_i \psi_0$ or $\mathcal{W}, w, \alpha \models_i \psi_1$.

5. if $\phi = \psi_0 \to \psi_1$, then
$$\mathcal{W}, w, \alpha \models_i \psi_0 \to \psi_1$$
iff for every v such that wRv, if $\mathcal{W}, w, \alpha \models_i \psi_0$, then $\mathcal{W}, w, \alpha \models_i \psi_1$.

6. if $\phi = \exists x\psi$, for some variable x, then
$$\mathcal{W}, w, \alpha \models_i \exists x\psi$$
iff there exists $a \in A_w$ such that $\mathcal{W}, w, \alpha[x \mapsto a] \models_i \psi$.

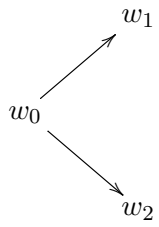
7. if $\phi = \forall x\psi$, for some variable x, then
$$\mathcal{W}, w, \alpha \models_i \forall x\psi$$
iff for all v such that wRv and all $a \in A_v$, $\mathcal{W}, w, \alpha[x \mapsto a] \models_i \psi$

As usual, if ϕ is a sentence, then we define
$$\mathcal{W}, w \models_i \phi$$
iff for every valuation α
$$\mathcal{W}, w, \alpha \models_i \phi$$

Convention 57. If $\mathcal{W} = \langle W, R, \{\mathcal{A}_w\}_{w \in W}\rangle$ is a Kripke model for First-Order Intuitionistic Logic, then for every $w \in W$ the universe of $\mathcal{A}_w$ will be denoted with A_w and the interpretation function with τ_{A_w}.

Remark 58. Let us observe that, for every Kripke model $\mathcal{W} = \langle W, R, \{\mathcal{A}_w\}_{w \in W}\rangle$ and every valuation α
$$\mathcal{W}, w\alpha[x \mapsto a] \models_i P(x)$$
iff $a \in \tau_w(P)$, where $\mathcal{A}_w = \langle A_w, \tau_w\rangle$.

Example 59. Let $W = \{w_0, w_1, w_2\}$ and $R = \{\langle w_0, w_1\rangle, \langle w_0, w_2\rangle, \langle w_0, w_0\rangle, \langle w_1, w_1\rangle, \langle w_2, w_2\rangle\}$. Hence W and R might be depicted



Let us work over signature with one unary predicate P. Define $A_{w_0} = \{a, b\}$, $A_{w_1} = \{a, b, c\}$, $A_{w_2} = \{a, b, d\}$ and

$$\begin{aligned} \tau_{w_0}(P) &= \{a, b\} \\ \tau_{w_1}(P) &= \{a, b, c\} \\ \tau_{w_2}(P) &= \{a, b\} \end{aligned} \tag{1}$$

Finally let $\mathcal{A}_{w_0} = \langle A_{w_0}, \tau_{w_0}\rangle$, $\mathcal{A}_{w_1} = \langle A_{w_1}, \tau_{w_1}\rangle$, $\mathcal{A}_{w_2} = \langle A_{w_2}, \tau_{w_2}\rangle$, and

$$\mathcal{W} = \langle W, R, \{\mathcal{A}_w\}_{w\in W}\rangle.$$

Then $\mathcal{W}, w_0 \nVdash_i \forall x P(x)$. Indeed let α be any valuation. Then, unravelling the definition we have:

$$\mathcal{W}, w_0, \alpha \models_i \forall x P(x)$$

iff

$$\text{for all } v \text{ such that } w_0 R v \text{ and all } a \in A_v \; \mathcal{W}, v, \alpha[x \mapsto a] \models_i P(x)$$

The above condition holds if and only if each of the three conditions below is fulfilled

1. $A_{w_0} = \tau_{w_0}(P)$
2. $A_{w_1} = \tau_{w_1}(P)$
3. $A_{w_2} = \tau_{w_2}(P)$

The last condition however does not hold, since $d \in A_{w_2} \setminus \tau_{w_2}(P)$.

Exercise 17. Let M be the following Kripke model for intuitionistic logic. M has infinitely many worlds w_n for $n \in \mathbb{N}$ (here $0 \in \mathbb{N}$). For each n, the domain of w_n is $\mathbb{N}$. A model w_i is accessible from w_j if and only if $i \geq j$. The signature consists of one unary predicate $P(x)$ such that in the world w_n, $P(k)$ holds iff $k \leq n$. Check whether the following sentences are satisfied in M:

- $\exists x\ P(x)$.
- $\neg\forall x\neg P(x)$.
- $\neg\exists x\neg P(x)$.
- $\forall x\exists y\ (P(x) \to \neg P(y))$.
- $\exists x(P(x) \to \forall y\ P(y))$.

Exercise 18. Let M be the following Kripke model for the intuitionistic logic. M has infinitely many worlds w_n for $n \in \mathbb{N}$. For each n, the domain of w_n is $\{0, 1, \ldots, n\}$. A model w_i is accessible from w_j if and only if $i \geq j$. The signature consists of one unary predicate $P(x)$ such that in the world w_n, $P(k)$ holds iff $k < n$ (note that in each world there is exactly one element x which does not satisfy $P(x)$). Check whether the following sentences are satisfied in M:

- $\neg\forall x P(x)$.
- $\exists x\neg P(x)$.
- $\forall x\exists y\ \neg P(y)$.
- $\exists x(\exists y P(y) \to P(x)))$.

Exercise 19. Find a sentence which is true in exactly one of the models described in the above problems.

Exercise 20. For each of the following formulae, find a model in which they are not satisfied.

- $\neg\neg\forall x P(x) \to \exists x P(x)$.
- $\forall x\exists y(P(x) \vee \neg P(y))$.
- $\exists x\forall y(P(x) \to P(y))$.

Observe that all the above sentences are classically valid.

Exercise 21. Find a model in which only one of the following sentences holds:

- $\neg P(a) \vee Q(a)$.
- $P(a) \to Q(a)$.

Exercise 22. Find a model in which only one of the following sentences holds:

- $\forall x(\exists y Q(y) \vee P(x))$.
- $\exists y\, Q(y) \vee \forall x\, P(x)$

5 Incompleteness

5.1 Peano Arithmetic

Peano Arithmetic is theory canonical formal theory of natural numbers. As we will see, it can be also viewed as a theory of finite sets. We will first investigate it as a theory of numbers, i.e. try to understand its basic properties.

Definition 60. Peano Arithmetic (PA) is a theory over the language with symbols $(S, +, \times, 0)$, called LPA, given by the following axioms:

1. $\forall x\big(S(x) \neq 0\big)$.
2. $\forall x\big(S(x) = S(y) \to x = y\big)$.
3. $\forall x\big(x + 0 = x\big)$.
4. $\forall x, y\big(x + S(y) = S(x + y)\big)$.
5. $\forall x\big(x \times 0 = 0\big)$.
6. $\forall x, y\big(x \times S(y) = x \times y + x\big)$.

Ind $\forall y_1, \ldots, y_n\Big[\phi(0, y_1, \ldots, y_n) \wedge \forall x\Big(\phi(x, y_1, \ldots, y_n) \to \phi(S(x), y_1, \ldots, y_n)\Big) \to \forall x \phi(x, y_1, \ldots, y_n)\Big]$,

where $\phi(x, y_1, \ldots, y_n) \in \mathscr{L}_{PA}$ is ans arbitrary formula.

Let us comment upon the above definition. PA comprises six axioms which are intended to define the basic operations and an infinite family of axioms—an axiom scheme—called the Induction Scheme (Ind). Thus PA actually has infinitely many axioms. Basically, the induction scheme is a poor's man version of saying that any object can be obtained via applying the successor function S to the number 0. This cannot be spelled out in a

proper manner in first-order logic and the induction axioms are intended to at least partially capture this intuition.

It might seem that axioms of PA as they stand are way too weak to capture any interesting facts of natural numbers. However, induction scheme gives us an incredible boost in strength.

Now we will state and proof a couple of facts about the arithmetic of natural numbers in PA. When writing "PA" next to the head of a theorem of proposition, we mean that it is provable within PA.

Proposition 61 (PA)**.** *For every x, either $x = 0$ or there exists y such that $x = S(y)$.*

Proof. We prove by induction on x that either $x = 0$ or there exists y such that $x = S(y)$.

If $x = 0$, then it clearly holds. So suppose that the claim holds for a given x. Then it trivially holds for $S(x)$. Therefore, for all x either $x = 0$ or there exists y such that $x = S(y)$. □

Proposition 62 (PA)**.** *Addition is associative, i.e. for all x, y, z,*

$$x + (y + z) = (x + y) + z.$$

Proof. Fix arbitrary x_0 and y_0. We prove by induction on z that for all z, $x_0 + (y_0 + z) = (x_0 + y_0) + z$.

If $z = 0$, then $y_0 + z = 0$, so $x_0 + (y_0 + z) = x_0 + y_0 = (x_0 + y_0) + 0 = (x_0 + y_0) + z$.

Suppose that the claim holds for z, we shall prove it for $S(z)$. We have the following equations:

$$\begin{aligned} x_0 + (y_0 + S(z)) &= x_0 + S(y_0 + z) \\ &= S(x_0 + (y_0 + z)) \\ &= S((x_0 + y_0) + z) \\ &= (x_0 + y_0) + S(0). \end{aligned}$$

The third equality follows by induction hypothesis, the rest follow by the axioms for addition.

By induction we can conclude that for all z

$$x_0 + (y_0 + z) = (x_0 + y_0) + z.$$

Since x_0, y_0 were arbitrary, we conclude that addition is associative. □

Proposition 63 (PA). *Addition is commutative, i.e., for any x, y*

$$(x + y) = (y + x).$$

In order to prove the proposition, we shall need one lemma.

Lemma 64 (PA). *For all x, $0 + x = x$.*

Proof. Let us prove the claim by induction on x. The claim is clear for $x = 0$, since then $0 + x = 0 + 0 = 0$, by addition axioms. So suppose that the claim holds for x and let us prove it for $S(x)$. We have as follows:

$$\begin{aligned} 0 + S(x) &= S(0 + x) \\ &= S(x). \end{aligned}$$

The first equality holds by axioms for addition and the second by induction hypothesis. The claim follows by induction. □

Lemma 65 (PA). *For all x, $S(0) + x = x + S(0)$.*

Proof. This time, we shall not prove the claim by induction. Take arbitrary x. Then the following holds:

$$\begin{aligned} S(0) + x &= S(0 + x) \\ &= S(x) \\ &= S(x + 0) \\ &= x + S(0). \end{aligned}$$

The second equality follows by Lemma 64. □

of Proposition 63. Fix any y_0. We will prove by induction on x that for all x

$$x + y_0 = y_0 + x.$$

For $x = 0$, the claim has been proved as Lemma 64. So suppose that the claim holds for a fixed x and let us prove it for $S(x)$. We have the following:

$$\begin{aligned} S(x) + y_0 &= S(x + 0) + y_0 \\ &= (x + S(0)) + y_0 \\ &= x + (S(0) + y_0) \\ &= x + (y_0 + S(0)) \\ &= (x + y) + S(0) \\ &= (y + x) + S(0) \\ &= y + (x + S(0)) \\ &= y + S(x). \end{aligned}$$

Above, we used associativity of addition and Lemma 65. By induction, we conclude that for all x, $x + y_0 = y_0 + x$. Since y_0 was arbitrary, we conclude that $+$ is commutative. □

Now, we will introduce one of the very basic notions of arithmetic.

Definition 66. By $x < y$ we mean $\exists z\ (z \neq 0 \wedge y = x + z)$. By $x \leq y$ we mean $x = y \vee x < y$.

Proposition 67 (PA)**.** *For all x, y exactly one of the following holds:*

1. $x = y$.
2. $x < y$.
3. $y < x$.

Proposition 68 (PA)**.** $\leq$ *is a partial order, i.e.*

1. $\leq$ *is transitive.*
2. $\leq$ *is reflexive.*
3. $\leq$ *is weakly-antisymmetric.*

Proposition 69 (PA)**.** 0 *is the least element in the order $\leq$. For any element x, $S(x)$ is the least element which is greater than x.*

It is clear by definition of order and associativity and commutativity of addition that for all a, b, c, d if $a \leq b$ and $c \leq d$, then

$$a + c \leq b + d.$$

We can also prove the following proposition:

Proposition 70 (PA)**.** *For all a, b, c, d if $a \leq b$ and $c \leq d$, then*

$$a \cdot c \leq b \cdot d.$$

Exercise 23. Prove in PA that multiplication is associative.

Exercise 24. Prove in PA that multiplication is commutative.

Exercise 25. Prove in PA that for all x, y, either $x = y$ or $x < y$, or $y < x$.

Exercise 26. Prove in PA that for all x, y, z, if $x \leq y$ and $y \leq z$, then $x \leq z$.

5.2 Arithmetic as set theory

The previous subsection was devoted to developing basic properties of arithmetic as number theory. It turned out that PA handles such properties as commutativity or associativity of addition. Now, we will show that PA is capable of formalising a good part of set theory. First, we will state one fact which is the key to formalising set theory in PA.

Theorem 71. *There exists a formula* $\exp(x, y)$ *in the language* $\mathscr{L}_{PA}$ *such that* PA *proves the following:*

1. $\exp(0, 1)$.
2. $\exp(S(x), y) \equiv \exists z \Big(\exp(x, z) \wedge y = (SS0) \cdot x \Big)$.

Intuitively, $\exp(x, y)$ holds if and only if $2^x = y$. One can observe that the theorem gives the value of 2^0 and explains how the value of 2^{x+1} depends on the value of 2^x. Therefore, it follows the same pattern as the axioms for addition and multiplication. Once we know how a given function behaves at 0 and how it values changes at successor steps, we intuitively should be able to recover the whole possible information on the behaviour of the given function by induction. Note that we can use induction with respect to formulae containing $\exp(x, y)$, since it is given an arithmetical formula. Let us state some arithmetical features of the exponentiation.

Now, we can give the central definition of this subsection.

Definition 72. We define $x \in y$ as the abbreviation of the following formula: by

$$\exists a, r \Big(y = a \cdot 2^{Sx} + 2^x + r \wedge r < 2^x \Big).$$

First, let us note that the definition of $x \in y$ should be stated more accurately as:

$$\exists a, r, u, w \Big(y = a \cdot u + \cdot w + r \wedge r < w \wedge \exp(x, w) \wedge \exp(S(x), u) \Big).$$

I.e., we should eliminate the expression 2^x, since there is no such term in the language. However, any formula containing it can be actually rewritten as above, so that the problematic expression is no more used.

Let us comment what is the motivation behind the definition of $x \in y$. Our intention is that the elements about which PA speaks are numbers. Actual natural numbers have binary expansions. This means that any number

can be viewed as a unique sequence of zeroes and ones. $x \in y$ means that the x-th bit of the number y in the binary expansion is 1.

In order to work with this notion, we shall need a bunch of basic facts about the arithmetic of exponentiation and, consequently, about the relation $x \in y$.

Proposition 73 (PA). *For all x, $x < 2^x$.*

Proposition 74 (PA). *For all x, y,*

$$2^x \cdot 2^y = 2^{x+y}.$$

Proposition 75 (PA). *For all x, y, if $x \leq y$, then*

$$2^x \leq 2^y.$$

Now we will show how to recover the basic features of set-theoretic $x \in y$ in PA.

Proposition 76 (PA). *There exists x such that for all y, $y \notin x$.*

Proof. We claim that 0 has no elements. Suppose that $y \in 0$. Then $y < 0$ which cannot hold by Proposition 69. □

Following the usual language of set theory, we shall use symbol $\emptyset$ to denote 0, whenever we will think of 0 as a set.

Proposition 77 (PA). *Let x, y be any two elements. Then there exists z such that for all u, $u \in z$ iff $u = x$ or $u = y$.*

Proof. Take any x, y and let $z = 2^x + 2^y$ if $x \neq y$ and $z = 2^x$ if $x = y$. Then $x \in z$ and $y \in z$ by Propositions 74 and 75. □

Proposition 78 (PA). *Take any x, y. If for all z, $z \in x \equiv z \in y$, then $x = y$.*

Proof. We prove the following claim by induction on b: for all $x, y \leq b$ if for all z, $z \in x$ iff $z \in y$, then $x = y$. If the claim holds for all b, then clearly the proposition holds as well.

For $b = 0$, the claim holds by Proposition 69. So suppose that the claim holds for b, and let us show that it holds for $S(b)$.

Take any $x, y \leq S(b)$ such that for all z, $z \in x$ iff $z \in y$. If $x, y < S(b)$, then the claim holds by induction hypothesis. So assume that exactly one of x, y is strictly smaller than $S(b)$. Without loss of generality we can assume that $x < S(b) = y$.

Now, either y is even or odd. Note that there exists y' such that $y = 2y'$ exactly if $0 \notin y$.

If y is odd, i.e. $0 \in y$, then we also have $0 \in x$. Then $x - 1, y - 1$ are both $\leq S(b)$ and have exactly the same elements. Therefore x, y also have the same elements, since all their elements are either elements x' or 0.

If $y = 2y'$, then note that for all z, $z \in y'$ iff $S(z) \in y$. Since $0 \notin y$, we also have $0 \notin x$, and therefore $x = 2x'$ for some x'. Furthermore, for all z, $z \in x'$ iff $S(z) \in x$. Note that x', y' have the same elements, so by induction hypothesis, there are equal. So x, y are equal as well. □

Proposition 79. *For every formula $\phi(u, y_1, \ldots, y_n)$* PA *proves the following: for every $a_1, \ldots, a_n$ and every b there exists x such that for every $u \leq b$, $u \in x$ iff $\phi(u, a_1, \ldots, a_n)$.*

Proof. We prove the claim for a fixed ϕ by induction on b. Fix any $a_1, \ldots, a_n$. If b equals 0, then the claim is obvious. Suppose that we have proved the claim for a given b and let us prove it for $S(b)$.

Let x' be such that for all $u \leq b$, $u \in x'$ iff $\phi(u, a_1, \ldots, a_n)$. Such x' exists by induction hypothesis. Now, if $\phi(S(b), a_1, \ldots, a_n)$, then let

$$x = x' + 2^{S(b)}$$

and let $x = x'$ otherwise. We claim that such an x satisfies our condition. Take any $u \leq S(b)$. Either $u \leq b$ or $u = S(b)$. If $u \leq b$, then $u \in x$ iff $u \in x'$ iff $\phi(u, a_1, \ldots, a_n)$ by induction hypothesis and the definition of x. If $u = S(b)$, then $u \in x$ iff $\phi(u, a_1, \ldots, a_n)$ directly by the definition of x. □

Proposition 80. *For any arithmetical formula $\phi(u, v_1, \ldots, v_n)$,* PA *proves the following: for all $a_1, \ldots, a_n$ and for all x, there exists y such that for all u,*

$$u \in y \equiv \phi(u, a_1, \ldots, a_n) \wedge u \in x.$$

Note that the above proposition is the Comprehension scheme for the arithmetical formula $\in$.

Proof. Fix any formula ϕ. We will prove that PA proves the instance of the comprehension scheme for ϕ. Fix arbitrary $a_1, \ldots, a_n$ and an arbitrary x. Let y be the set of $u \leq x$ such that $\phi(u, a_1, \ldots, a_n)$ and $u \in x$.

Note that such a y exists by Proposition 79 applied to the formula $\phi(u, a_1, \ldots, a_n) \wedge u \in x$ (This might possibly be a subtle point). Then y is exactly the set of u such that $\phi(u, a_1, \ldots, a_n)$ and $u \in x$, since by definition of $u \in x$ and Proposition 73, if $u \in x$, then $u < x$. □

Proposition 81 (PA). *For all x, there exists y such that for all z, $z \in y$, exactly if $z \subseteq y$.*

In the formulation of the above proposition, $z \subseteq y$ is defined in the familiar fashion: $z \subseteq y$, if for any u, if $u \in z$, then $u \in y$. Thus, the proposition states that for each x, there exists a powerset of x, which we denote with the familiar notation $\mathscr{P}(x)$.

Exercise 27. Write down all x such that $x \in n$ for $n = 1$, $n = 17$, $n = 30$, $n = 64$, $n = 100$, $n = 215$.

In the following exercises, you may freely use Theorem 71 and the results stated in these notes (provided that the claim of the exercise isn't itself proved in these notes, in which case, you can use all results which occurred prior to it).

At this point, be somewhat careful about what you assume. Note that, for example, we still do not know whether sets have any well defined *number of element* (in fact, they do).

Exercise 28. Prove that for all x, $x < 2^x$.

Exercise 29. Prove that for all x, y, $2^{x+y} = 2^x 2^y$.

Exercise 30. Prove that for all x, y, if $y \subseteq x$, then $y \leq x$.

Exercise 31. Prove that for all x, there exists $\mathscr{P}(x)$.

We have almost shown that PA proves all axioms of the set theory ZFC for the arithmetical elementhood predicate $x \in y$. The only thing, we still have to check is that it actually proves the separation scheme. We begin with a related fact with a slightly more arithmetical flavour.

Theorem 82. *For all arithmetical formulae $\phi(x, y, z_1, \ldots, z_k)$,* PA *proves as follows:*

$$\forall z_1, \ldots, z_k \, \forall a \Big(\forall x < a \exists y \, \phi(x, y, z_1, \ldots, z_k) \rightarrow \exists b \forall x < a \exists y < b \, \phi(x, y, z_1, \ldots, z_k) \Big)$$

The scheme of the form given in the theorem is called **Collection Scheme**. Let us comment upon it: the scheme expresses that if for each x from the initial segment $\{0, 1, \ldots, a\}$, we can choose some y such that $\phi(x, y)$ holds, then actually there is some common bound b such that this y may be required to be smaller than b. The intuitive proof goes as follows: for each x smaller than a, let $y(x)$ be some chosen element y such that $\phi(x, y)$. Then

there only a many elements $y(0), y(1), \ldots, y(a-1)$, so there is some element greater than all of them. Thus, in a way, the Collection Scheme expresses the same intuition as the Induction Scheme: every element may be reached from 0 in finitely many steps. Now, let us prove the theorem.

Proof. Fix any formula $\phi(x, y, z_1, \ldots, z_k)$. We shall prove in PA the instance of the induction scheme for the formula ϕ. Fix any $z_1, \ldots, z_k$. By induction on a we prove that if for $x < a$ there exists y such that $\phi(x, y, z_1, \ldots, z_k)$, then there exists b such that for all $x < a$ there exists $y < b$ such that $\phi(x, y, z_1, \ldots, z_k)$. For $a = 0$, the claim is obvious (it holds vacuously). So let us assume that the claim is true for a given a and let us prove it for $a+1$. So suppose that for all $x < a+1$ there exists a y such that $\phi(x, y)$. Fix any y_0 such that $\phi(a, y_0)$. By induction hypothesis, we know that there exists b such that for all $x < a$, there exists $y < b$ such that $\phi(x, y)$. Let $b' = \max(b, y_0 + 1)$. Then for every $x < a + 1$ (i.e. every x such that either $x < a$ or $x = a$), there exists $y < b'$ such that $\phi(x, y)$, since we can take either $y < b$ or $y = y_0$. □

Using the collection scheme one can prove Separation Scheme.

Fact 83. *For any formula $\phi(x, y, z_1, \ldots, z_k)$,* PA *proves as follows:*

$$\forall z_1, \ldots, z_k \forall A \Big(\text{Func}_\phi(A) \to \exists B \, \forall x \in A \exists y \in B \; \phi(x, y) \Big),$$

where $\text{Func}_\phi(A)$ *is the abbreviation of the following formula:*

$$\text{Func}_\phi(A) = \forall x \in A \exists y \; \phi(x, y) \wedge \forall x, y', y'' \Big(\phi(x, y') \wedge \phi(x, y'') \to y' = y'' \Big).$$

Notice that $\text{Func}_\phi(A)$ says that the formula $\phi(x, y)$ defines a function with the domain A, i.e., for every $x \in A$ it ascribes exactly one y.

Now that we know that arithmetic may be treated as a set theory, we can reconstruct a good part of set theory within PA. Let us introduce some set-theoretic notions which will be of great use in the further development of the theory of syntax within PA.

Definition 84. By the **ordered pair** of elements a, b we mean the set $\{a, \{b\}\}$. We denote it by $\langle a, b \rangle$.

Fact 85. *For arbitrary a, b, c, d, $\langle a, b \rangle = \langle c, d \rangle$ if and only if $a = c$ and $b = d$.*

Definition 86. By a **product** of A, B, denoted $A \times B$, we mean the set of elements of the form $\langle a, b \rangle$ where $a \in A, b \in B$.

Definition 87. We say that R is a **relation** with the domain A and the codomain B, when it is a subset of $A \times B$.

In the above definition, the intended reading of $\langle a, b\rangle \in R$ is that the relation R holds between elements a and b. So to represent a relation we just list all the pairs of elements which are linked by this very relation.

Definition 88. We say that f is a **function** with the domain A and the codomain B, denoted $f : A \to B$, when $f \subseteq A \times B$ and for all $a \in A$ there exists $b \in B$ such that $\langle a, b\rangle \in f$.

If $\langle a, b\rangle \in f$, we denote this fact by "$f(a) = b$."

When we write $\langle a, b\rangle \in f$, we really just mean that the value of the element f at the argument a is b. Similarly to the case of relations, in order to represent a function, we just list pairs of the form: $\langle$ argument, values $\rangle$.

Note that all the above definitions are exactly the same as in the usual set theory. We literally copy those definitions in PA, as we know how strongly the properties of the elementhood relation in arithmetic resemble those of elementhood relation in set theory.

Definition 89. We say that s is a **sequence** of length n if it is a function with the domain $\{1, 2, \ldots, n\}$. We denote the length of s with $\mathrm{lh}(s)$.

Typically, one defines the domain of a sequence as $\{0, 1, \ldots, n\}$. We decided that the above version would be more convenient to use. We denote a sequence whose only elements are $a_1, a_2, \ldots, a_n$ with

$$\langle a_1, a_2, \ldots, a_n\rangle.$$

In particular, by $\langle a\rangle$, we mean a sequence of length one whose only terms is a.

Definition 90. Let s, t be sequences. By a **concatenation** of s and t, $s \frown t$, we mean the sequence of length $\mathrm{lh}(s) + \mathrm{lh}(t)$ such that for all $i \leq \mathrm{lh}(s \frown t)$,

$$s \frown t(i) = \begin{cases} s(i), & \text{if } i \leq \mathrm{lh}(s) \\ t(j), & \text{if } i > \mathrm{lh}(s) \text{ and } j = i - \mathrm{lh}(s). \end{cases}$$

Note that the concatenation of sequences s and t is just the sequence s followed by the sequence t. E.g.,

$$\langle 1, 2, 6, 1\rangle \frown \langle 2017, 1, 3\rangle = \langle 1, 2, 6, 1, 2017, 1, 3\rangle.$$

Once we have the notion of sequence in PA, we can imitate recursive definitions. Let us begin with a discussion of a concrete example.

Definition 91. We say that m is n **factorial**, denoted $n!$, if there exists a sequence s of length n such that $s(1) = 1$ and for all $i < n$

$$s(i+1) = s(i) \cdot (i+1),$$

and $m = s(n)$.

Note that in the above definition, the sequence s keeps track of how we define the factorial of the number $n+1$ assuming we know the factorial of n. Then the definition of $n!$ states that it is the product of this process after n steps. Let us see one more example in this spirit.

Definition 92. For arbitrary a, b, we say that $m = a^b$ if there exists a sequence s of length $b+1$ such that $s(1) = 1$ and for all $i \leq b$,

$$s(i+1) = s(i) \cdot a,$$

and $s(b+1) = m$.

Note that in the above definition $s(i+1) = a^i$. The shift in the definition may be possibly confusing.

5.3 Arithmetisation of syntax

Now we get really close to Gödel's theorem. We shall show how PA can handle syntactic notions such as terms, formulae or proofs. To begin with, we have to say what do we mean by a primitive character. In order to do this, we just pick some arbitrary elements which we will stipulate to be fixed characters from the language of PA. The only thing we have to take care of, is that those elements will not turn out to be any more complex structures, like terms, formulae, etc. Since all the syntactic objects will be defined as functions, it is enough to guarantee that primitive symbols will not be sequences at all.

Convention 93. We choose the following numbers as the name of the characters from the language $\mathscr{L}_{PA}$:

- $\ulcorner 0 \urcorner = \{\emptyset\}$.
- $\ulcorner S \urcorner = \{\{\emptyset\}\}$.
- $\ulcorner + \urcorner = \{\{\{\emptyset\}\}\}$.
- $\ulcorner \times \urcorner = \underbrace{\{\ldots\{}_{4 \text{ times}}\emptyset\}\ldots\}$.

- $\ulcorner(\urcorner = \underbrace{\{\ldots\{}_{5 \text{ times}} \emptyset\}\ldots\}$.
- $\ulcorner)\urcorner = \underbrace{\{\ldots\{}_{6 \text{ times}} \emptyset\}\ldots\}$.
- $\ulcorner\neg\urcorner = \underbrace{\{\ldots\{}_{7 \text{ times}} \emptyset\}\ldots\}$.
- $\ulcorner\wedge\urcorner = \underbrace{\{\ldots\{}_{8 \text{ times}} \emptyset\}\ldots\}$.
- $\ulcorner\vee\urcorner = \underbrace{\{\ldots\{}_{9 \text{ times}} \emptyset\}\ldots\}$.
- $\ulcorner\rightarrow\urcorner = \underbrace{\{\ldots\{}_{10 \text{ times}} \emptyset\}\ldots\}$.
- $\ulcorner\forall\urcorner = \underbrace{\{\ldots\{}_{11 \text{ times}} \emptyset\}\ldots\}$.
- $\ulcorner\exists\urcorner = \underbrace{\{\ldots\{}_{12 \text{ times}} \emptyset\}\ldots\}$.
- $\ulcorner=\urcorner = \underbrace{\{\ldots\{}_{13 \text{ times}} \emptyset\}\ldots\}$.

Additionally, we say that x is the code of l-th variable, denoted $\mathrm{Var}(x, l)$ if x is of the form $\underbrace{\{\ldots\{}_{14+2l \text{ times}} \emptyset\}\ldots\}$.

Note that in the above convention e.g. by $\{\{\emptyset\}\}$, we mean the only number x such that there exists only one y with $y \in x$, and moreover, this y also only has one element which is the empty set. One can check that such an x is actually equal to 2^{2^0}. In order to define what we mean by iterating this construction $2l + 14$ times, we have to refer to sequences, as in the definition of factorial or exponentiation.

Now we are ready to give the first definition of a syntactic object in PA.

Definition 94 (PA)**.** We say that τ is a **term** of $\mathscr{L}_{PA}$ if there exists a sequence s of length l such that $s(l) = \tau$ and for all $i \leq l$ one of the following conditions holds:

- $s(i) = \ulcorner 0 \urcorner$.

- There exist m, v such that $s(i) = \langle v \rangle$ and $\mathrm{Var}(v, m)$, i.e., $s(i)$ is a sequence of length one whose only element is the code of the variable v_m.
- There exists $j < i$ such that

$$s(i) = \langle S, \ulcorner(\urcorner \rangle \frown s(j) \frown \langle \ulcorner)\urcorner \rangle.$$

- There exist $j, k < i$ such that

$$s(i) = \langle \ulcorner(\urcorner \rangle \frown s(j) \frown \langle \ulcorner+\urcorner \rangle \frown s(k) \frown \langle \ulcorner)\urcorner \rangle.$$

- There exist $j, k < i$ such that

$$s(i) = \langle \ulcorner(\urcorner \rangle \frown s(j) \frown \langle \ulcorner\times\urcorner \rangle \frown s(k) \frown \langle \ulcorner)\urcorner \rangle.$$

Let us comment on the above definition: we want to define arithmetical terms so that they are sequences of some specified numbers which we agreed to treat as primitive characters. We want this sequence to satisfy certain syntactic condition so that it is not equal to, e.g., $\langle \ulcorner\forall\urcorner, \ulcorner\forall\urcorner, \ulcorner 0\urcorner, \ulcorner(\urcorner \rangle$ or some other meaningless expression.

The definition of correctly built term is inductive: 0 is a correctly built terms and if s, t are correctly built terms, then $s + t$ is a correctly built term as well. The general way to handle inductive definitions in PA is to say that there exists a sequence which keeps track of all the steps of building up some given expression. This is precisely what the definition of the terms says. The sequence s from the above definition is called the **generating sequence** of the term τ.

Definition 95 (PA)**.** We say that ϕ is a **formula** of $\mathscr{L}_{PA}$ if there exists a sequence s of length l such that $s(l) = \phi$ and for all $i \leq l$ one of the following conditions holds:

- There exist terms σ, τ such that $s(i) = \sigma \frown \langle \ulcorner=\urcorner \rangle \tau$.
- There exists $j \leq i$ such that $s(i) = \langle \ulcorner\neg\urcorner, \ulcorner(\urcorner \rangle \frown s(j) \frown \langle \ulcorner)\urcorner \rangle$.
- There exist $j, k \leq i$ such that $s(i) = \langle \ulcorner(\urcorner \rangle \frown s(j) \frown \langle \ulcorner\wedge\urcorner \rangle \frown s(k) \frown \langle \ulcorner)\urcorner \rangle$.
- There exist $j, k \leq i$ such that $s(i) = \langle \ulcorner(\urcorner \rangle \frown s(j) \frown \langle \ulcorner\vee\urcorner \rangle \frown s(k) \frown \langle \ulcorner)\urcorner \rangle$.

- There exist $j, k \leq i$ such that $s(i) = \langle \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner \rightarrow \urcorner \rangle \frown s(k) \frown \langle \ulcorner) \urcorner \rangle$.
- There exists $j \leq i$ such that $s(i) = \langle \ulcorner \exists \urcorner, v, \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner) \urcorner \rangle$ and v is a code of the k-th variable for some $k \leq \phi$.
- There exists $j \leq i$ such that $s(i) = \langle \ulcorner \forall \urcorner, v, \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner) \urcorner \rangle$ and v is a code of the k-th variable for some $k \leq \phi$.

Again, we call s the **generating sequence** of ϕ. Note that this sequence is in fact highly non-unique, since it may comprise some formulae which are not even subformulae of ϕ.

Now, we will define some syntactic notions whose definition is mildly more complicated. In definitions of terms and formulae we basically had to track the build-up of these object in order to ensure that there are generated using only some constructions from a restricted list. Now, we will have to deal in construction in which we make some additional bookkeeping along the way as we construct our objects.

Definition 96. Let τ be a term of $\mathscr{L}_{PA}$. We call x the **set of free variables** of τ if there exist sequences s, s' of the same length l such that s is the generating sequence of τ, $s'(l) = x$, and the following conditions are satisfied for all $i \leq l$:

- If $s(i) = v$ is a variable, then $s'(i) = \{v\}$.
- If $s(i) = 0$, then $s'(i) = \emptyset$.
- If $s(i) = \langle \ulcorner S \urcorner, \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner) \urcorner \rangle$ for some $j < i$, then $s'(i) = s(i)$.
- If $s(i) = \langle \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner + \urcorner \rangle \frown s(k) \frown \langle \ulcorner) \urcorner \rangle$ for some $j, k < i$, then $s'(i) = s'(j) \cup s'(k)$.
- If $s(i) = \langle \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner + \urcorner \rangle \frown s(k) \frown \langle \ulcorner) \urcorner \rangle$ for some $j, k < i$, then $s'(i) = s'(j) \cup s'(k)$.

Let ϕ be a formula from $\mathscr{L}_{PA}$. We call x the **set of free variables** of ϕ if there exist sequences s, s' of the same length l such that s is the generating sequence of ϕ, $s'(l) = x$, and the following conditions are satisfied for all $i \leq l$:

- If $s(i) = \sigma \frown \langle \ulcorner = \urcorner \rangle \tau$ for some terms σ, τ, then $s'(i) = a \cup b$, where a is the set of free variables of σ and b is the set of free variables of τ.

- If $s(i) = \langle \ulcorner \neg \urcorner, \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner) \urcorner \rangle$ for some $j < i$, then $s'(i) = s'(j)$.
- If $s(i) = \langle \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner \vee \urcorner \rangle \frown s(k) \frown \langle \ulcorner) \urcorner \rangle$ or $s(i) = \langle \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner \vee \urcorner \rangle \frown s(k) \frown \langle \ulcorner) \urcorner \rangle$ for some $j, k < i$, then $s'(i) = s'(j) \cup s'(k)$.
- If $s(i) = \langle \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner \to \urcorner \rangle \frown s(k) \frown \langle \ulcorner) \urcorner \rangle$ or $s(i) = \langle \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner \to \urcorner \rangle \frown s(k) \frown \langle \ulcorner) \urcorner \rangle$ for some $j, k < i$, then $s'(i) = s'(j) \cup s'(k)$.
- If $s(i) = \langle \ulcorner \exists \urcorner, v, \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner) \urcorner \rangle$ or $s(i) = \langle \ulcorner \forall \urcorner, v, \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner) \urcorner \rangle$ for some $j \leq i$, then $s'(i) = s'(j) \setminus \{v\}$.

We denote the set of free variables of a term τ or a formula ϕ by $\mathrm{FV}(\tau), \mathrm{FV}(\phi)$, respectively.

The idea behind a definition of the set of free variables is again very simple: we know how to define inductively what a correctly built term or a formula is. We may then define its set of free variables by going through this construction and taking side notes on the set of free variables of the currently examined subformula or subterm of the given formula or term. In a similar manner, we may construct the substitution function.

Definition 97. Let τ, t be terms of $\mathscr{L}_{PA}$ and let v be any variable. We say that τ' is the effect of **substitution** of the term t in the term τ for the variable v if there exist sequences s, s' of the same length l such that s is the generating sequence of τ and for all $i \leq l$ the following conditions are satisfied:

- If $s(i) = \langle \ulcorner 0 \urcorner \rangle$, then $s'(i) = \langle \ulcorner 0 \urcorner \rangle$.
- If $s(i) = \langle \ulcorner w \urcorner \rangle$ for some variable w, then $s'(i) = t$, if $w = v$ and $s'(i) = \langle \ulcorner w \urcorner \rangle$, if $w \neq v$.
- If there exists $j < i$ such that $s(i) = \langle S, \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner) \urcorner \rangle$, then
$$s'(i) = \langle S, \ulcorner (\urcorner \rangle \frown s'(j) \frown \langle \ulcorner) \urcorner \rangle.$$
- If there exist $j, k < i$ such that $s(i) = \langle \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner + \urcorner \rangle \frown s(k) \frown \langle \ulcorner) \urcorner \rangle$, then
$$s'(i) = \langle \ulcorner (\urcorner \rangle \frown s'(j) \frown \langle \ulcorner + \urcorner \rangle \frown s'(k) \frown \langle \ulcorner) \urcorner \rangle.$$
- If there exist $j, k < i$ such that $s(i) = \langle \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner \times \urcorner \rangle \frown s(k) \frown \langle \ulcorner) \urcorner \rangle$, then
$$s'(i) = \langle \ulcorner (\urcorner \rangle \frown s'(j) \frown \langle \ulcorner \times \urcorner \rangle \frown s'(k) \frown \langle \ulcorner) \urcorner \rangle.$$

Suppose that ϕ is a formula from the language $\mathscr{L}_{PA}$. We say that ϕ' is the effect of **substitution** of a term t in a formula ϕ for the variable v if there exist sequences s, s' of the same length l such that s is the generating sequence of ϕ and for all $i \leq l$ the following conditions hold:

- If there exist terms σ, τ such that $s(i) = \sigma \frown \langle \ulcorner = \urcorner \rangle \tau$, then $s'(i) = \sigma' \frown \langle \ulcorner = \urcorner \rangle \tau'$ where σ', τ' result via substituting the term t for the variable v in σ, τ, respectively.

- If there exists $j \leq i$ such that $s(i) = \langle \ulcorner \neg \urcorner, \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner) \urcorner \rangle$, then

$$s'(i) = \langle \ulcorner \neg \urcorner, \ulcorner (\urcorner \rangle \frown s'(j) \frown \langle \ulcorner) \urcorner \rangle$$

- If there exist $j, k \leq i$ such that $s(i) = \langle \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner \wedge \urcorner \rangle \frown s(k) \frown \langle \ulcorner) \urcorner \rangle$, then

$$s'(i) = \langle \ulcorner (\urcorner \rangle \frown s'(j) \frown \langle \ulcorner \wedge \urcorner \rangle \frown s'(k) \frown \langle \ulcorner) \urcorner \rangle.$$

- If there exist $j, k \leq i$ such that $s(i) = \langle \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner \vee \urcorner \rangle \frown s(k) \frown \langle \ulcorner) \urcorner \rangle$, then

$$s'(i) = \langle \ulcorner (\urcorner \rangle \frown s'(j) \frown \langle \ulcorner \vee \urcorner \rangle \frown s'(k) \frown \langle \ulcorner) \urcorner \rangle.$$

- If there exist $j, k \leq i$ such that $s(i) = \langle \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner \rightarrow \urcorner \rangle \frown s(k) \frown \langle \ulcorner) \urcorner \rangle$, then

$$s'(i) = \langle \ulcorner (\urcorner \rangle \frown s'(j) \frown \langle \ulcorner \rightarrow \urcorner \rangle \frown s'(k) \frown \langle \ulcorner) \urcorner \rangle.$$

- If there exists $j \leq i$ such that $s(i) = \langle \ulcorner \exists \urcorner, w, \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner) \urcorner \rangle$ and w is a code of the k-th variable for some $k \leq \phi$, then

$$s'(i) = \langle \ulcorner \exists \urcorner, w, \ulcorner (\urcorner \rangle \frown s'(j) \frown \langle \ulcorner) \urcorner \rangle$$

  if $v = w$ or else

$$s'(i) = s(i).$$

- If there exists $j \leq i$ such that $s(i) = \langle \ulcorner \forall \urcorner, w, \ulcorner (\urcorner \rangle \frown s(j) \frown \langle \ulcorner) \urcorner \rangle$ and w is a code of the k-th variable for some $k \leq \phi$, then

$$s'(i) = \langle \ulcorner \forall \urcorner, w, \ulcorner (\urcorner \rangle \frown s'(j) \frown \langle \ulcorner) \urcorner \rangle$$

  if $v = w$ or else

$$s'(i) = s(i).$$

If ϕ' is the effect of substitution of a term t in the formula ϕ for the variable v, we denote ϕ' with Subst(ϕ, t, v) or $\phi[t/v]$.

Note that we again define the effect of substitution of a term t in a formula ϕ for a variable v by modifying the generating sequence of ϕ: we replace the occurrences of v with t. However, note that if the variable v happens to be bounded by a quantifier we forget the corresponding formula resulting by substitution and replace it with the currently examined subformula of ϕ.

6 Self-reference and Gödel's Theorem

In this section, we will finally present a proof of the celebrated Gödel's Incompleteness Theorem. We will first state it in a very specialised version which does not yet give the right impression of how general phenomenon it describes.

Theorem 98 (Gödel's Incompleteness Theorem)**.** *There exists a sentence γ in the language $\mathscr{L}_{PA}$ such that* PA $\nvdash \gamma$ *and* PA $\nvdash \neg\gamma$*.*

Let us first give a sketch of the proof. It will be neither formal, nor particularly accurate (to put it mildly), but should give some understanding of the motivating ideas.

So let γ be a sentence saying "I am not provable in PA." Suppose γ has a proof in PA. Then the sentence γ is not provable in PA. So PA is inconsistent. On the other hand, if the negation of γ is provable in PA, this implies that γ has a proof in PA. This again would yield PA inconsistent. So if PA is consistent than it cannot either prove or refute γ.

As we have said, the above sketch is not really an accurate description of the proof. What is more, it is not at all clear how to translate this sketch to the actual proof.

To begin with, let us show in what sense can a sentence speak anything of itself. This relies on a very clever trick.

Theorem 99 (Gödel's Diagonal Lemma)**.** *Let $\phi(x)$ be any formula from $\mathscr{L}_{PA}$ with one free variable. Then there exists a sentence γ such that*

$$Q \vdash \gamma \equiv \phi(\ulcorner\gamma\urcorner).$$

If a sentence γ is a string of characters $s_1, \ldots, s_n$, then by $\ulcorner\gamma\urcorner$ we mean the number

$$\langle \ulcorner s_1 \urcorner, \ldots, \ulcorner s_n \urcorner \rangle.$$

This notation will be used also in the case of other syntactic objects. So: if x is a string of symbols, then by $\ulcorner x \urcorner$ we mean the number defined as the sequence of codes of these symbols.

Recall that Q is Robinson Arithmetic, the theory whose axioms are that of PA minus the induction scheme. Before we proceed to the proof let us introduce a bit of notation.

Definition 100 (PA)**.** We say that y is the **numeral** for the number x, $N(x, y)$, if there exists a sequence s of length $x+1$ such that $N(1) = 0$, $N(x+1) = y$, and for any $i < x + 1$,

$$N(i + 1) = \langle \ulcorner S \urcorner, (\rangle \frown N(i) \frown \langle) \rangle.$$

In other words, the numeral for x is the code of the canonical term denoting the number x. We will use $N(x)$ to denote y as if it were a term. This abbreviation may be clearly eliminated from any formula.

Definition 101 (PA)**.** Let ψ be a formula with only one free variable v (i.e. $\mathrm{FV}(\phi) = \{v\}$ for some $v \in \mathrm{Var}$. Then by

$$\mathsf{Subst}(\phi, t),$$

we mean

$$\mathsf{Subst}(\phi, t, v).$$

We denote it also simply with $\phi(t)$.

Now we are ready to prove the Diagonal Lemma.

Proof of the Diagonal Lemma. Fix any formula $\phi(x)$ from $\mathscr{L}_{PA}$ with one free variable. Let $\delta(x, y)$ be the formula

$$y = \mathsf{Subst}(x, N(x)).$$

Let

$$\psi(x) = \exists y\big(\delta(x, y) \wedge \phi(y)\big).$$

Let

$$a = \ulcorner \psi \urcorner.$$

Let

$$\gamma = \psi(\underline{a})$$

where $\underline{a}$ denotes

$$\underbrace{S \ldots S}_{a\text{times}} 0.$$

Then the following equalities and equivalences hold:

$$\begin{aligned}
\gamma &= \exists y\big(\delta(\underline{a}, y) \wedge \phi(y)\big) \\
&= \exists y\big(y = \mathsf{Subst}(a, N(a)) \wedge \phi(y)\big) \\
&= \exists y\big(y = \mathsf{Subst}(\psi, N(a)) \wedge \phi(y)\big) \\
&\equiv \exists y\big(y = \ulcorner \psi(\underline{a}) \urcorner \wedge \phi(y)\big) \\
&\equiv \phi(\ulcorner \gamma \urcorner).
\end{aligned}$$

All the equalities are obtained simply by expanding definitions. The last equivalence is by pure logic and the definition of γ, the previous one is obtained by executing the substitution in the definition of y. □

The proof of the Diagonal Lemma may leave a false impression that it is based on some intricacies of coding. However, we really use only basic properties of substitution. The sentence γ rewritten in the natural language, reads roughly as follows:

> The effect of substitution for the only free variable in the formula
>
> *The effect of substitution for the only free variable in the formula x the numeral denoting the formula x satisfies ϕ.*
>
> the numeral denoting the formula
>
> *The effect of substitution for the only free variable in the formula x the numeral denoting the formula x satisfies ϕ.*
>
> satisfies ϕ.

The reader is encouraged to check what happens if one actually computes the effect of the substitution in question.

We are almost ready to prove Gödel's Incompleteness Theorem. We almost know how to produce a sentence which says of itself that it is unprovable. We only have to make sure that we can say that something is provable.

Definition 102 (PA)**.** Let ϕ be a formula. We say that it is an **axiom of** PA ($\mathrm{Ax}_{\mathrm{PA}}(\phi)$ or $\phi \in \mathrm{Ax}_{\mathrm{PA}}$) if one of the following conditions hold:

1. $\phi = \ulcorner \forall x\big(S(x) \neq 0\big) \urcorner$.
2. $\phi = \ulcorner \forall x\big(S(x) = S(y) \to x = y\big) \urcorner$.
3. $\phi = \ulcorner \forall x\big(x + 0 = x\big) \urcorner$.
4. $\phi = \ulcorner \forall x, y\big(x + S(y) = S(x + y)\big) \urcorner$.

5. $\phi = \ulcorner \forall x \big(x \times 0 = 0 \big) \urcorner$.

6. $\phi = \ulcorner \forall x, y \big(x \times S(y) = x \times y + x \big) \urcorner$.

7. There exist formulae ψ, ξ and a variable v such that

$$\xi = \langle \ulcorner (\urcorner \rangle \frown \psi[0/v] \frown \langle \ulcorner \wedge \urcorner, \ulcorner \forall \urcorner, \ulcorner x \urcorner, \ulcorner (\urcorner \rangle \frown \psi[x/v] \frown \langle \ulcorner \rightarrow \urcorner \rangle \frown \psi[S(x)/v] \frown \langle \ulcorner) \urcorner, \ulcorner) \urcorner, \ulcorner \rightarrow \urcorner, \ulcorner \forall \urcorner, \ulcorner x \urcorner, \ulcorner (\urcorner \rangle \frown \psi[x/v] \frown \langle \ulcorner) \urcorner \rangle.$$

   and there exists a sequence s of length l such that $s(1) = \xi, s(l) = \phi$, and for every $i < l$ there exists a variable y_i such that

$$s(i+1) = \langle \ulcorner \forall \urcorner, y_i, \ulcorner (\urcorner \rangle \frown s(i) \frown \langle \ulcorner) \urcorner \rangle.$$

Note that the above condition simply formalises what shape a formula must have to count as an axiom of PA: in particular, the last clause says what does it mean to be an induction axiom. For clarity, we actually have omitted several parentheses in definition of ξ.

Let us separate a useful notion from the above definition:

Definition 103 (PA)**.** Let ϕ be a sentence and ψ a formula. We say that ϕ is a **universal closure** of ψ if there exists a sequence s of length l such that $s(1) = \psi$, $s(l) = \phi$, and for every $i < l$ there exists a variable y such that

$$s(i+1) = \langle \ulcorner \forall \urcorner, y, \ulcorner (\urcorner \rangle \frown s(i) \frown \langle \ulcorner) \urcorner \rangle.$$

Now we will define the notion of proof in PA. A slightly awkward thing is that we actually have not defined what a proof in first-order logic is in the first place. We hope that the reader will believe us that we can define some axiom schemes which together with the *modus ponens* rule suffice to axiomatise entailment in first-order logic. We will now write down these axioms already in their formalised version.

In order to facilitate reading, from now on we will say, e.g., that a formula has the shape

$$\psi \wedge \psi$$

rather than

$$\langle \ulcorner (\urcorner \rangle \psi \frown \langle \wedge \rangle \frown \psi \frown \langle \ulcorner) \urcorner \rangle,$$

since it shouldn't lead to any confusion.

Definition 104 (PA)**.** We say that d is a **proof** of a formula ϕ in PA ($\mathrm{Prf}(d, \phi)$) if d is a sequence of length l such that for every $i \leq l$ at least one of the following conditions is satisfied:

1. $d(i)$ is a universal closure of an axiom of PA.
2. There exists a formula ψ such that $d(i)$ is a universal closure of the formula $\psi \to \psi$.
3. There exist formulae ψ, η such that $d(i)$ is a universal closure of the formula $\psi \to (\eta \to \psi)$.
4. There exist formula ψ, η, ξ such that $d(i)$ is a universal closure of the formula $\psi \to (\eta \to \xi) \to ((\psi \to \eta) \to (\psi \to \xi))$.
5. There exist formulae ψ, ξ such that $d(i)$ is a universal closure of the formula $(\neg\psi \to \neg\xi) \to (\xi \to \psi)$.
6. There exist a formula ψ, a variable v, and a term t such that $d(i)$ is a universal closure of the formula $\forall v(\psi) \to \psi[t/v]$.
7. There exist formulae ψ, η, and a variable v such that $d(i)$ is a universal closure of the formula $\forall v(\psi \to \eta) \to (\forall v(\psi) \to \forall v(\eta))$.
8. There exist a formula ψ and a variable $v \notin \mathrm{FV}(\psi)$ such that $d(i)$ is a universal closure of the formula $\psi \to \forall v(\psi)$.
9. There exists a variable v such that $d(i)$ is a universal closure of the formula $v = v$.
10. There exist a formula ψ and variables u, v, w such that $d(i)$ is a universal closure of the formula $u = v \to (\psi[u/w] \to \psi[v/w])$.
11. There exist formulae ψ, η such that $d(i)$ is a universal closure of the formula $(\psi \vee \eta) \to (\neg\psi \to \eta)$.
12. There exist formulae ψ, η such that $d(i)$ is a universal closure of the formula $(\neg\psi \to \eta) \to (\psi \vee \eta)$.
13. There exist formulae ψ, η such that $d(i)$ is a universal closure of the formula $(\psi \wedge \eta) \to \neg(\psi \to \neg\eta)$.
14. There exist formulae ψ, η such that $d(i)$ is a universal closure of the formula $\neg(\psi \to \neg\eta) \to (\psi \wedge \eta)$.
15. There exist a formula ψ and a variable v such that $d(i)$ is a universal closure of the formula $\exists v(\psi) \to \neg\forall v(\neg\psi)$.
16. There exist a formula ψ and a variable v such that $d(i)$ is a universal closure of the formula $\neg\forall v(\neg\psi) \to \exists v(\psi)$.

17. There exist $j, k < i$ such that for some formulae ψ, η $d(j)$ is of the form $\psi \to \eta$, $d(k) = \psi$ and $d(i) = \eta$.

Notes that the above definition says in effect that proof is a sequence of formulae in which every formula is either an axiom of PA, an axiom of the first-order logic or results from the previous steps in the proof by applying *modus ponens*.

Definition 105 (PA)**.** Let ϕ be a formula. We say that it is **provable** if there exists a proof d of ϕ. Then we write $\mathrm{Pr}(\phi)$.

Now, we can proceed to Gödel's Theorem.

Sketch of the proof 98. Let γ be a sentence such that:

$$\mathrm{PA} \vdash \gamma \equiv \neg \mathrm{Pr}(\ulcorner \gamma \urcorner).$$

That such a sentence exists is guaranteed by Theorem 99. We claim that PA proves neither γ nor its negation.

Suppose PA proves γ. Then there exists a proof $\psi_1, \ldots, \psi_l$ of the sentence γ from the axioms of PA. Let

$$d = \langle \ulcorner \psi_1 \urcorner, \ldots, \ulcorner \psi_l \urcorner \rangle.$$

One can check by induction on l that in such case $\mathrm{PA} \vdash \mathrm{Prf}(d, \ulcorner \psi_l \urcorner)$. This means that $\mathrm{PA} \vdash \mathrm{Pr}(\ulcorner \gamma \urcorner)$, so $\mathrm{PA} \vdash \neg\gamma$. This means that PA is inconsistent.

Suppose that $\mathrm{PA} \vdash \neg\gamma$. Then $\mathrm{PA} \vdash \exists d \mathrm{Prf}(d, \ulcorner gamma \urcorner)$. Notice that $\mathbb{N}$ with standard operations is a model of PA, therefore there exists $d \in \mathbb{N}$ such that $\mathrm{Prf}(d, \ulcorner \gamma \urcorner)$. We can check by induction on d' that if $\mathrm{Prf}(d', \ulcorner \phi \urcorner)$ holds in $\mathbb{N}$, then there actually exists a proof of ϕ in PA (d is in $\mathbb{N}$, so it makes sense to reason by induction). This means that there exists a proof of γ in PA. But this again means that PA is inconsistent.

□

Now, we will discuss a bunch of results related to Gödel's Theorem. We will begin with Gödel's second theorem which states that PA does not prove its own consistency. We will try, however, to state result with the right generality. On the other hand, we will be somewhat sloppy until the end of this section.

Definition 106. Let Th be a theory in a language with finitely many symbols containing the arithmetical symbols. We say that a formula $P(x)$ satisfies **provability conditions** if:

1. Th $\vdash \phi$ implies Th $\vdash P(\ulcorner\phi\urcorner)$.
2. Th $\vdash P(\ulcorner\phi\urcorner) \to P(\ulcorner P(\underline{\ulcorner\phi\urcorner})\urcorner)$.
3. Th $\vdash \forall\phi,\psi\ \Big(P(\phi\to\psi)\wedge P(\phi)\to P(\psi)\Big)$.

The last condition may be rewritten more (and still not fully) accurately as: Th $\vdash \forall x,y,z\ \Big((z = x \frown \langle\ulcorner\to\urcorner\rangle \frown y \wedge P(z)\wedge P(x))\to P(y)\Big)$.

Note that we haven't yet defined what the expression $\ulcorner\phi\urcorner$ means at all when ϕ is not a string of characters from the arithmetical language. However, for languages extending the language of PA with finitely many new symbols we can define coding of its syntax in a fairly straightforward manner.

Theorem 107. *Suppose that* Th *is a theory extending Robinson Arithmetic Q in a language which extends the language of arithmetic with finitely many symbols and that* Th *proves provability conditions for a formula $P(x)$. Then*

$$\text{Th} \nvdash \neg P(\ulcorner\neg 0 = 0\urcorner).$$

Note that P satisfying provability conditions should be thought of as some form of a truth predicate. Then Gödel's second theorem states that no such theory can possibly prove its own consistency.

Proof. Let γ be such that

$$\text{Th} \vdash \gamma \equiv \neg P(\ulcorner\gamma\urcorner).$$

(Note that to prove Diagonal Lemma we really did not use any induction whatsoever. It was a matter of finite computation on actual codes.) Then

$$\text{Th} \vdash P(\ulcorner\gamma \equiv \neg P(\underline{\ulcorner\gamma\urcorner})\urcorner).$$

So

$$\text{Th} \vdash P(\ulcorner\gamma\urcorner) \equiv P(\ulcorner\neg P(\underline{\ulcorner\gamma\urcorner})\urcorner).$$

Now, by provability conditions we have:

$$\text{Th} \vdash P(\ulcorner\gamma\urcorner) \to P(\ulcorner P(\underline{\ulcorner\gamma\urcorner})\urcorner).$$

On the other, we already know that:

$$\text{Th} \vdash P(\ulcorner\gamma\urcorner) \to P(\ulcorner\neg P(\underline{\ulcorner\gamma\urcorner})\urcorner).$$

Obviously:

$$\text{Th} \vdash \gamma \wedge \neg\gamma \rightarrow \neg(0 = 0).$$

So applying provability conditions we obtain:

$$\text{Th} \vdash P(\ulcorner P(\underline{\ulcorner \gamma \urcorner})\urcorner) \wedge P(\ulcorner \neg P(\underline{\ulcorner \gamma \urcorner})\urcorner) \rightarrow P(\ulcorner 0 \neq 0 \urcorner).$$

So:

$$\text{Th} \vdash P(\ulcorner \gamma \urcorner) \rightarrow P(\ulcorner 0 \neq 0 \urcorner).$$

So suppose that $\text{Th} \vdash \neg P(\ulcorner 0 \neq 0 \urcorner)$. Then we obtain:

$$\begin{array}{lll} \text{Th} & \vdash & \neg P(\ulcorner \gamma \urcorner) \\ \text{Th} & \vdash & \gamma \\ \text{Th} & \vdash & P(\ulcorner \gamma \urcorner). \end{array}$$

Thus we have shown that Th is inconsistent.

□

Note that the above proof looks very much like formalised proof of Gödels first theorem. The conclusion of the latter proof was that $\text{PA} \nvdash \gamma$. Here it has been formalised within Th as:

$$\text{Th} \vdash \neg P(\ulcorner \gamma \urcorner).$$

However, now we have one extra "level" to our disposal and we can just say that $\text{Th} \vdash \gamma$. This allows us to draw the conclusion of the theorem.

Proposition 108. *The formula* $\text{Pr}(x)$ *satisfies provability conditions in* PA.

Sketch of the proof. 1. Suppose that $\text{PA} \vdash \phi$. Then there exists a proof $\phi_1, \ldots, \phi_n$ of ϕ. Let $d = \langle \ulcorner \phi_1 \urcorner, \ldots, \ulcorner \phi_n \urcorner \rangle$. Then we check by induction on d that $\text{PA} \vdash \text{Prf}(d, \ulcorner \phi \urcorner)$ (we have to check by induction all kinds of auxiliary facts, e.g. that if ψ is a formula with $\ulcorner \psi \urcorner \leq d$, then $\text{PA} \vdash \text{Form}(\ulcorner \psi \urcorner)$). 2. Since the argument in 1. used only some finite operations on syntactic symbols and induction, we can formalise it inside PA to conclude $\text{Pr}(\ulcorner \text{Pr}(\underline{\ulcorner \phi \urcorner})\urcorner)$ from the assumption that $\text{Pr}(\ulcorner \phi \urcorner)$.

3. We reason inside PA. If $\text{Prf}(d_1, \phi \rightarrow \psi)$ and $\text{Prf}(d_2, \phi)$, then $\text{Prf}(d_1 \frown d_2 \frown \psi, \psi)$ which we can check by definition of Prf. □

After Gödel's theorem was published, a question has been raised whether we can prove another seemingly "paradoxical" sentence "This sentence is provable in PA." It turned out however, that this latter sentence in fact *is* provable in PA and actually a much more general result holds.

Theorem 109 (Löb's Theorem)**.** *Take any* ϕ *be such that* PA $\vdash \mathrm{Pr}(\ulcorner\phi\urcorner) \to \phi$*. Then* PA $\vdash \phi$*.*

Proof. Using theorem 99, fix any β such that

$$\mathrm{PA} \vdash \beta \equiv (\mathrm{Pr}(\ulcorner\beta\urcorner) \to \phi).$$

Then

$$\mathrm{PA} \vdash \mathrm{Pr}(\ulcorner\beta\urcorner) \equiv (\mathrm{Pr}(\ulcorner\mathrm{Pr}(\underline{\ulcorner\beta\urcorner})\urcorner) \to \mathrm{Pr}\phi).$$

By provability conditions,

$$\mathrm{PA} \vdash \mathrm{Pr}(\beta) \to \mathrm{Pr}(\ulcorner\mathrm{Pr}(\underline{\ulcorner\beta\urcorner})\urcorner).$$

So if PA $\vdash \mathrm{Pr}(\ulcorner\phi\urcorner) \to \phi$, then

$$\mathrm{PA} \vdash \mathrm{Pr}(\ulcorner\beta\urcorner) \to \phi.$$

But by the equivalence satisfied by β, this implies:

$$\mathrm{PA} \vdash \beta.$$

By provability conditions,

$$\mathrm{PA} \vdash \mathrm{Pr}(\ulcorner\beta\urcorner).$$

But since PA proves both β and $\mathrm{Pr}(\ulcorner\beta\urcorner)$, this implies by the property of β:

$$\mathrm{PA} \vdash \phi.$$

□

In the proof of First Gödel's Theorem, we have used the fact that the set of natural numbers forms a model for PA. This may possibly make Gödel's Theorem less philosophically interesting. In order to prove that PA is incomplete, we have to assume that it is true about natural numbers. Let us now present a theorem which gets rid of this assumption. We will prove the theorem in case of PA which may seem strange but this only due to the fact that at this point we do not want to state the result in proper generality.

Theorem 110 (Rosser's Theorem)**.** *Suppose that* PA *is consistent. Then there exists a sentence* ρ *such that* PA *neither proves* ρ *nor refutes it.*

Proof. Let $\mathrm{Rprf}(d, \phi)$ be defined as

$$\mathrm{Prf}(d, \phi) \wedge \forall d' < d \, \neg \mathrm{Rprf}(d', \neg\phi).$$

Let $\mathrm{RPr}(\phi)$ be defined as $\exists d \; \mathrm{Rprf}(d, \phi)$. We call the formula RPr **Rosser's provability predicate**.

Let ρ be a formula such that

$$\mathrm{PA} \vdash \rho \equiv \neg \mathrm{RPr}(\ulcorner \rho \urcorner).$$

Now, if $\mathrm{PA} \vdash \rho$, then there exists a proof $\phi_1, \ldots, \phi_n$ of ρ. Let $d = \langle \ulcorner \phi_1 \urcorner, \ldots, \ulcorner \phi_n \urcorner \rangle$. Then we can check by induction on d that

$$\mathrm{PA} \vdash \mathrm{Prf}(d, \ulcorner \phi_n \urcorner).$$

Moreover,

$$\mathrm{PA} \vdash \forall x \Big(x \leq d \to (x = 0 \vee x = 1 \vee \ldots \vee x = d) \Big).$$

But since PA is consistent none of the $0, \ldots, d$ can be a code of actual proof of $\neg\rho$. Therefore,

$$\mathrm{PA} \vdash \forall \mathrm{RPr}(\ulcorner \rho \urcorner).$$

But then

$$\mathrm{PA} \vdash \neg\rho$$

which shows that PA is inconsistent.

Suppose now that

$$\mathrm{PA} \vdash \neg\rho$$

i.e.

$$\mathrm{PA} \vdash \mathrm{Rprf}(\ulcorner \rho \urcorner).$$

Let d be code of the proof of $\neg\rho$. Since PA is consistent, none of $0, 1 \ldots, d-1$ is a code of an actual proof of ρ. Thus PA proves that the least proof of a sentence from the pair $\rho, \neg\rho$ is actually a proof of $\neg\rho$, which contradicts $\mathrm{Rprf}(\ulcorner \rho \urcorner)$ and yields PA inconsistent. □

Rosser's theorem really should be proved in greater generality, since this is it were it becomes most interesting. Occasionally, one is interested in theories in the arithmetical language which are not true in $\mathbb{N}$. Then Rosser's theorem shows us that those theories still have undecidable sentences. Arguments in the vein of Rosser's theorem generally play a prominent role in the study of metamathematics, so by no means is it an isolated trick.