



Chapter 6

Modal Logic

H.C.M. (Harrie) de Swart

Abstract Modal operators, like ‘it is necessary that’ or ‘John knows that’, express an attitude about the proposition to which they are applied. Modal logic studies the reasoning in modal contexts, extending classical logic in which only connectives and quantifiers are taken into account. There are many systems of modal logic, depending on the axioms one wants to accept for the modal operators. The semantics of the modal operators is in terms of possible worlds, where each possible world is supposed to satisfy classical logic. A proposition is necessarily true if it is true in every world accessible or imaginable from the given world. Also tableau rules are available for the different systems of modal logic. Constructing a tableau-deduction in modal propositional logic of a formula from given premisses, if it exists, is straightforward; and if it does not exist, one easily constructs a counterexample from a failed attempt to construct one. Epistemic logic is about the modal operator ‘knowing that’ and an interesting puzzle in this field is the one of the muddy children. The possible world semantics is useful to understand a number of phenomena in the philosophy of language: rigid designators and the ‘de dicto - de re’ distinction. Also strict implication and counterfactuals may be understood in terms of possible world semantics. In modal predicate logic we study the behavior of modal operators in combination with the quantifiers. We shall see that in order to make sense, modal contexts should be referentially transparent and at the same time extensionally opaque.

6.1 Modal Operators

Although *modal operators* seldom occur in scientific proza, they do occur in daily language: it is *possible* that Rotterdam is the capital of Holland; it is *impossible* that living creatures can survive fire; it is *necessary* that each object is equal to itself; John *knows* that Amsterdam is the capital of the Netherlands; Rhea *believes* that his wife is the best there is; it is *obligatory* to stop for a red traffic light; it is (not) *permitted* to have a gun; John will *always* love Janet.

A *modal operator* expresses an attitude about the proposition to which it is applied. One distinguishes *alethic* operators, such as ‘it is necessary that’ and ‘it is possible that’, *epistemic* operators, such as ‘agent *i* knows that’ and ‘agent *i* believes that’, *deontic* operators, such as ‘it is obligatory that’ and ‘it is permitted that’ and *temporal* operators, such as ‘it is and always will be true that’.

In modal logic one studies reasoning in modal texts, i.e., texts which contain modal operators; see, for instance, Exercise 6.1. One may distinguish:

- *modal propositional logic*: it studies the reasoning in texts containing not only the classical propositional connectives, denoted by \Leftrightarrow , \rightarrow , \wedge , \vee and \neg , but also the modal operators of necessity, denoted by \Box , and possibility, denoted by \Diamond ; and
- *modal predicate logic*: it studies the reasoning in texts which in addition contain the quantifiers \forall and \exists .

Frege’s view in Section 4 of his *Begriffsschrift* [13] is that the notions of necessity and possibility belong to epistemology and involve a covert reference to human knowledge for which there is no place in pure logic.

C.I. Lewis’ book *A Survey of Symbolic Logic* [25] from 1918 is generally considered to be the beginning of modern modal logic. Rejecting material implication as an adequate representation of ‘if ..., then ...’, C.I. Lewis put forward a logic of *strict implication*, in which the latter can be rendered in terms of necessity and material implication: $\Box(A \rightarrow B)$.

For a brief outline of the history of modal logic we refer the reader to the *Historical Introduction* of E.J. Lemmon [24], pp. 1-12. For Aristotle’s modal logic and Megarian and Stoic Theories of Modality see W. & M. Kneale [20], pp. 81-96 and pp. 117-128 respectively.

In his *Reference and Modality*, W.V. Quine [32] argues that modal logic is problematic, because \Box , to be read as ‘it is necessary that’, and \Diamond , to be read as ‘it is possible that’, create a context for which Leibniz’ Law does not seem to hold. His argument is as follows: let $a = 9$ and $b =$ the number of planets. Then $a = b$. But $\Box(9 > 7)$ is considered to be true, while \Box (the number of planets > 7) is generally considered to be false; the number of planets might have been five, if it had pleased the Creator. So substitution of ‘the number of planets’ for ‘9’ in $\Box(9 > 7)$ turns a truth into a falsehood, while the number of planets = 9. However, this argument is misleading, since the expression 9 refers to a natural number, while the expression ‘the number of planets’ is a function that assigns to every possible world a natural number. And the number 9 cannot be equal to the function ‘the number of planets’. What is true is that $9 =$ the number of planets in this world, that $\Box(9 > 7)$ and that we hence also have to accept that \Box (the number of planets in this world > 7), which is not counter-intuitive at all. We shall come back to this issue in Subsection 6.6.2 and in Subsection 6.11.1.

Since the principle of extensionality (Leibniz’ Law) at first sight does not seem to hold for contexts containing modal, epistemic or psychological operators, such contexts have come to be called non-extensional or *intensional*. See L. Linsky [28]. For a closer investigation of this issue see Subsection 6.11.1 on Modal Predicate Logic and Essentialism.

A model-theoretic description of modal logic in terms of *possible worlds* was developed, in particular by S.A. Kripke in his paper *Semantical Analysis of Modal Logic* [21]. The basic idea may be said to be to treat modal contexts as involving a reference to more than one possible world or possible state of affairs. $\Box A$ holds in world w iff A holds in all worlds which are accessible from w and $\Diamond A$ holds in world w iff there is some world accessible from w in which A holds.

Exercise 6.1. Translate the following argument in the language of modal propositional logic: If I want to succeed [S], then I should make many exercises [E]. If I want to make many exercises, then I should have a lot of free time [L]. It is impossible to have a lot of free time. Therefore, it is impossible to succeed.

6.2 Different systems of Modal Logic

Modal logic results from classical logic by adding one (or two) connectives to the language of classical logic:

\Box , to be read as: ‘it is necessary that’; or as: ‘it is obligatory that’; or as: ‘agent i knows that’, etc., and

\Diamond , to be read as: ‘it is possible that’; or as: ‘it is permitted that’, etc.

However, $\Diamond A \Leftrightarrow \neg\Box\neg A$ is generally accepted as an axiom schema. Alternatively, one may define $\Diamond A$ as $\neg\Box\neg A$: A is possible iff $\neg A$ is not necessary; and A is permitted iff $\neg A$ is not obligatory.

With \Box and \Diamond added as unary operators to the language of classical logic, $\Box P$, $\Diamond P$, $\Box P \rightarrow \Diamond P$, $\Box(\Box P \rightarrow \Diamond P)$, $\Box\Box P$, $\Box\Diamond P$, $\Diamond\Box P$, $\Diamond\Diamond P$, and so on, become formulas of our extended language. Using \Diamond we may translate the expression ‘ P is *contingent*’ by $\Diamond P \wedge \Diamond\neg P$; and the expression ‘ P is *compatible with* Q ’ as $\Diamond(P \wedge Q)$.

Since \Box may have different (alethic, deontic, epistemic, tense) readings or interpretations, it comes as no surprise that there are many different axioms one may postulate for \Box . Even the meaning of the word ‘necessary’ may vary:

- logically necessary, like in: ‘if I walk fast, then I walk fast’ is logically necessary;
- physically necessary, like in: it is physically necessary that if I drop this pencil, then it falls to the ground;
- ethically necessary, like in: ‘one should not kill’ is ethically necessary.

However, in general the notion of necessity is not a very clear one: ‘men are necessarily mortal’ may mean ‘all men are mortal’ or ‘from certain biological laws it follows that men are mortal’ or ‘from the history up till now it follows that men are mortal’; and the reader may discover other meanings as well.

Depending on the intended meaning of the modal operator \Box one may accept or reject one or more axioms for \Box . For instance, $\Box A \rightarrow A$ seems plausible for the alethic interpretation of \Box : if A is (logically or physically) necessary, then A will be the case; but the same formula is not plausible for the deontic reading of \Box : from A is obligatory, it does not have to follow that A is actually the case. On the other hand, the formula $\Box A \rightarrow \Diamond A$ seems plausible for the deontic reading of \Box : if A is obligatory, then A is permitted.

By imposing different conditions on \Box , many modal logics result. Below we list some of the more important systems of modal logic.

The modal logic K (named after Kripke) results from classical propositional logic by adding to the axioms of (classical) propositional logic for \rightarrow , \wedge , \vee and \neg (see Section 2.6) and the rule Modus Ponens (from A and $A \rightarrow B$ deduce B) one axiom schema and one rule of inference for \Box :

axiom schema: $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

rule: $\frac{\vdash A}{\vdash \Box A}$ i.e., if A is a theorem (of modal logic), then $\Box A$ is too.

The modal logic KT is obtained from K by adding the axiom schema

$$T: \Box A \rightarrow A.$$

The modal logic $S4 = KT4$ is obtained from KT by adding the axiom schema

$$4: \Box A \rightarrow \Box \Box A,$$

and the modal logic $S5 = KT4E$ is obtained from $KT4 = S4$ by adding the axiom schema

$$E: \Diamond A \rightarrow \Box \Diamond A.$$

Under the epistemic reading, the 4-axiom $\Box A \rightarrow \Box \Box A$ is called *positive introspection*: if I know A , then I know that I know A ; and the E -axiom $\Diamond A \rightarrow \Box \Diamond A$ is called *negative introspection*: if I do not know $\neg A$, then I know that I do not know $\neg A$.

Definition 6.1. By $K-$ we shall mean any of the systems K , KT , $KT4 = S4$, or $KT4E = S5$.

Definition 6.2. The alphabet of the language of modal propositional logic consists of the following symbols:

P_1, P_2, P_3, \dots , called propositional variables or atomic formulas;

the operators \Leftrightarrow , \rightarrow , \wedge , \vee , \neg and \Box ; and the brackets (and).

Definition 6.3 (Formulas of modal propositional logic).

P_1, P_2, P_3, \dots are formulas of modal propositional logic;

If A and B are formulas of modal propositional logic, then also $(A \Leftrightarrow B)$, $(A \rightarrow B)$, $(A \wedge B)$ and $(A \vee B)$ are formulas of modal propositional logic;

If A is a formula of modal propositional logic, then also $(\neg A)$ and $(\Box A)$ are formulas of modal propositional logic.

Definition 6.4. $\Diamond A := \neg \Box \neg A$.

Warning: $\Box \neg A$ or, equivalently, $\neg \Diamond A$ means ‘ $\neg A$ is necessary’ or, equivalently, A is impossible. Notice that in $\Box \neg A$ the negation concerns A . But $\neg \Box A$ or, equivalently, $\Diamond \neg A$ means ‘ A is not necessary’ or, equivalently, $\neg A$ is possible. Notice that in $\neg \Box A$ the negation concerns \Box .

Convention We can minimize the need for parentheses by agreeing that we leave out the most outer parentheses in a formula and that in

$$\Leftrightarrow, \rightarrow, \wedge, \vee, \neg, \Box$$

any connective has a higher rank than any connective to the right of it and a lower rank than any connective to the left of it.

According to this convention, $\Box A \wedge B \rightarrow C$ should be read as $((\Box A) \wedge B) \rightarrow C$, i.e., if A is necessary and (in addition) B , then C , because \rightarrow has a higher rank than \wedge and \wedge has a higher rank than \Box . This formula is different from the formula $(\Box(A \wedge B)) \rightarrow C$, i.e., if $A \wedge B$ is necessary, then C , and also different from the formula $\Box((A \wedge B) \rightarrow C)$, i.e., it is necessary that if $A \wedge B$, then C . According to our convention, the formula $\Box \neg A \vee B$ should be read as $(\Box \neg A) \vee B$, because \vee has a higher rank than \neg and \Box , and not as $\Box((\neg A) \vee B)$, nor as $\Box(\neg(A \vee B))$, which mean quite something else.

Definition 6.5 (Deduction; deducible). Let A_1, A_2, \dots, A_n and B be formulas of modal propositional logic. A *deduction* of B from A_1, A_2, \dots, A_n in the modal propositional logic $K-$ is a finite sequence of formulas with B as last one, such that each formula in this sequence is either one of the formulas A_1, A_2, \dots, A_n , or one of the logical axioms of $K-$, or is obtained by applying one of the rules to formula(s) earlier in the sequence.

B is *deducible* from A_1, A_2, \dots, A_n in $K-$ iff there exists a deduction of B from A_1, A_2, \dots, A_n in $K-$. **Notation:** $A_1, A_2, \dots, A_n \vdash B$ in $K-$.

In case $n = 0$, i.e., there are no premisses A_1, A_2, \dots, A_n , we say that B is *provable* in $K-$. **Notation:** $\vdash B$ in $K-$.

Example 6.1. $\vdash A \rightarrow \Diamond A$ in KT and also $\vdash \Box A \rightarrow \Diamond A$ in KT .

Proof. $\Box \neg A \rightarrow \neg A$ is an axiom of KT . Since $\Box \neg A \rightarrow \neg A \vdash A \rightarrow \neg \Box \neg A$ in classical propositional logic (contraposition), it follows that $\vdash A \rightarrow \Diamond A$ in KT . Both $\Box A \rightarrow A$ and $A \rightarrow \Diamond A$ are provable in KT and because $\Box A \rightarrow A$, $A \rightarrow \Diamond A \vdash \Box A \rightarrow \Diamond A$ in classical propositional logic, it follows that $\Box A \rightarrow \Diamond A$ is provable in KT . \square

Exercise 6.2. Show that a) $A \rightarrow \Box \Diamond A$ and b) $\neg \Box A \rightarrow \Box \neg \Box A$ are provable in $S5$.

Exercise 6.3 (Cosmological argument for God's existence).

Let P stand for 'something exists' and Q for 'there is a perfect being (God exists)'. Show that: $\Diamond P, \Box(\Diamond P \rightarrow Q) \vdash \Box Q$ in $S5$. [From Hubbeling [19], Section 8; 'cosmological' because of the occurrence of $\Diamond P$]

Exercise 6.4 (Ontological proof of God's existence). Let Q stand for 'God exists'. Show that: $\Box(Q \rightarrow \Box Q), \Diamond Q \vdash Q$ in $S5$. [This argument is Hartshorne's version of Anselm's ontological proof of God's existence (Anselm, Proslogion III); see Hubbeling [19], Section 8.]

Exercise 6.5. Find the mistake made in the following putative deduction in the modal logic $S5$ of Q (God exists) from $Q \rightarrow \Box Q$ and $\Diamond Q$.

1. $\Box Q \vee \neg \Box Q$
2. $\Box Q \vee \Box \neg \Box Q$ From 1 and exercise 6.2.
3. $\neg \Box Q \rightarrow \neg Q$ From the premiss $Q \rightarrow \Box Q$.
4. $\Box \neg \Box Q \rightarrow \Box \neg Q$ From 3 and the axioms and rule for \Box .
5. $\Box Q \vee \Box \neg Q$ From 2 and 4.
6. $\Box Q$ From 5 and the premiss $\Diamond Q$.
7. Q From 6 and $\Box Q \rightarrow Q$.

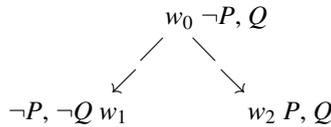
Exercise 6.6 (Ross's Paradox). Prove directly from the definitions:

i) $\vdash \Box A \rightarrow \Box(A \vee B)$ in K , and ii) $\vdash \Diamond A \rightarrow \Diamond(A \vee B)$ in K .

Notice that these theorems at first sight look counter-intuitive in the case of deontic logic, reading $\Box A$ as 'it is obligatory that A ' or ' A ought to be the case'. See, however, the discussion of deontic logic in Section 6.3.

6.3 Possible World Semantics

Clearly, the truth of $\Box A$ depends on more than just the truth value of A . We say that $\Box A$ is true in the present world/situation w iff A is true in all worlds/situations w' which are accessible/imaginable from w . And that $\Diamond A$ is true in world w iff there is a world w' accessible from w such that A is true in world w' . Consider, for instance, the following state of affairs: Jane is cleaning the street with water. So, in the present world/situation w_0 , it does not rain ($\neg P$) and the street becomes wet (Q). In the present world/situation, Jane can imagine two other possible worlds, one (w_1) in which it does not rain ($\neg P$) and the street does not become wet ($\neg Q$) and another one (w_2) in which it does rain (P) and the street becomes wet (Q). We may model this state of affairs with the following (Kripke) model M :

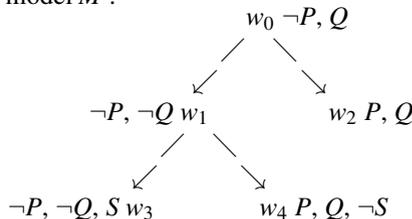


Given this state of affairs or Kripke model M , $\Box(P \rightarrow Q)$ (necessarily: if it rains, then the street becomes wet) is true in world w_0 because in every world Jane can imagine, i.e., in worlds w_0, w_1, w_2 , it is true that if it rains, then the street becomes wet, in other words, in all three worlds, $\neg P$ is true or Q is true.

And $\Diamond P$ (it is possible that it rains) is true in world w_0 , because Jane can imagine a world w' , namely w_2 , in which P is true.

We may describe this Kripke model M by the tuple $M = \langle \{w_0, w_1, w_2\}, R, \models \rangle$, where the accessibility relation R is defined by $w_0 R w_0$, $w_0 R w_1$ and $w_0 R w_2$, and where \models is defined by $w_0 \not\models P$, $w_0 \models Q$, $w_1 \not\models P$, $w_1 \not\models Q$, $w_2 \models P$ and $w_2 \models Q$. Clearly, the picture contains all this information.

Of course, in world (situation) w_1 Jane may imagine two other possible worlds (situations): w_3 , in which $\neg P$ and $\neg Q$ hold, and in addition the sun is shining (S), and w_4 , in which P , Q and $\neg S$ are true. This state of affairs is then described by the following Kripke model M' :



If in model M' it holds that world w_3 is accessible from world w_0 , i.e., w_0Rw_3 , then $\Diamond S$ (it is possible that the sun is shining) is true in world w_0 of model M' . However, if not w_0Rw_3 , then $\Diamond S$ is not true in world w_0 of model M' .

This brings us to the general definition of a Kripke model.

Definition 6.6 (Kripke model). $M = \langle W, R, \models \rangle$ is a Kripke model iff

- W is a non-empty set, the elements of which are called *possible worlds*;
- R is a binary relation on W , called the *accessibility relation*; wRw' is to be read as: world w' is accessible from world w ;
- \models is a relation between the elements of W and the atomic formulas; $w \models P$ is to be read as: atomic formula P is true in world w .

In the case of deontic logic, wRw' is read as: w' is a (deontically) perfect alternative of w .

Definition 6.7 ($M, w \models A$). Given a Kripke model $M = \langle W, R, \models \rangle$, we define $M, w \models A$ (to be read as: A is true (holds) in world w of model M) for arbitrary w in W and for arbitrary formulas A (of modal propositional logic) as follows:

- $M, w \models P := w \models P$ (P atomic).
- $M, w \models B \wedge C := M, w \models B$ and $M, w \models C$.
- $M, w \models B \vee C := M, w \models B$ or $M, w \models C$.
- $M, w \models B \rightarrow C :=$ not $M, w \models B$ or $M, w \models C$.
- $M, w \models \neg B :=$ not $M, w \models B$, also written as $M, w \not\models B$.
- $M, w \models \Box B :=$ for all w' in W , if wRw' , then $M, w' \models B$.
- $M, w \models \Diamond B :=$ there is a world w' in W such that wRw' and $M, w' \models B$.

Note that the connectives \wedge , \vee , \rightarrow and \neg in each world w are treated as in classical logic; in other words, classical logic applies in each possible world, i.e., a Kripke model can be conceived as a collection of classical models, supplemented by an accessibility relation.

Definition 6.8 ($M \models A$). Let $M = \langle W, R, \models \rangle$ be a Kripke model and A a formula.

M is a Kripke model of A (or A is true in M) := for every world w in W , $M, w \models A$.

Notation: $M \models A$. 'Not $M \models A$ ' is also denoted by: $M \not\models A$.

It is easy to check that the axiom for K , i.e., $\Box(B \rightarrow C) \rightarrow (\Box B \rightarrow \Box C)$, is true in every world w of every Kripke model M , i.e., for all Kripke models M , $M \models \Box(B \rightarrow C) \rightarrow (\Box B \rightarrow \Box C)$. We shall say that $\Box(B \rightarrow C) \rightarrow (\Box B \rightarrow \Box C)$ is *valid*.

Proof. Suppose $M, w \models \Box(B \rightarrow C)$, i.e., for all w' in M , if wRw' , then $M, w' \models B \rightarrow C$. (1)

Next, suppose $M, w \models \Box B$, i.e., for all w' in M , if wRw' , then $M, w' \models B$. (2)

Then it follows from (1) and (2) that for all worlds w' in M , if wRw' , then $M, w' \models C$, i.e., $M, w \models \Box C$. \square

Instead of saying that $\Box(B \rightarrow C) \rightarrow (\Box B \rightarrow \Box C)$ is valid, we may also say that $\Box B \rightarrow \Box C$ is a *valid consequence* of $\Box(B \rightarrow C)$.

Definition 6.9 (Valid consequence; valid). B is a *valid consequence* of premisses $A_1, \dots, A_n :=$ for all Kripke models M and for every world w in M , if $M, w \models A_1 \wedge \dots \wedge A_n$, then $M, w \models B$. **Notation:** $A_1, \dots, A_n \models B$. In case $n = 0$, i.e., there are no premisses, we say that B is *valid*, i.e., for all Kripke models M and for all worlds w in M , $M, w \models B$. **Notation:** $\models B$.

Notice that $A_1, \dots, A_n \models B$ iff $\models A_1 \wedge \dots \wedge A_n \rightarrow B$.

It is also easy to verify that the only rule for \Box (if $\vdash A$, then $\vdash \Box A$) preserves validity: if $\models A$, then $\models \Box A$.

Proof. Suppose that $\models A$, i.e., for all Kripke models M and for every world w in M , $M, w \models A$. (1)

We have to show that for all M and for all w in M , $M, w \models \Box A$, i.e., for all w' in M , if wRw' , then $M, w' \models A$. This follows trivially from (1). \square

So, we have shown the following theorem:

Theorem 6.1.

1. $\models \Box(B \rightarrow C) \rightarrow (\Box B \rightarrow \Box C)$; *equivalently:* $\Box(B \rightarrow C) \models (\Box B \rightarrow \Box C)$.
2. if $\models A$, then $\models \Box A$.

The \Box -axiom for KT , $\Box A \rightarrow A$, is not in all worlds of all Kripke models true. The following Kripke model $M = \langle \{w_0, w_1\}, R, \models \rangle$ with $w_0 R w_1$, but not $w_0 R w_0$, is a counterexample:



$M, w_0 \models \Box P$, but $M, w_0 \not\models P$. In world w_0 of this Kripke model M , P (stopping for a red traffic light) is obligatory, meaning that P is true in all deontically perfect alternatives of w_0 , but P does not have to be true in w_0 .

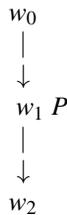
It is easy to see that $\Box A \rightarrow A$ holds precisely in those Kripke models $M = \langle W, R, \models \rangle$ in which the accessibility relation R is *reflexive*, i.e., for all w in M , wRw . For if $M, w \models \Box A$, and R is reflexive, then clearly $M, w \models A$.

Deontic logic If one reads $\Box A$ as 'it ought to be the case that A ' (or, equivalently, as 'A is obligatory') and $\Diamond A$ as 'it is permitted that A ' (or, equivalently, as 'A is permissible'), one speaks of *deontic logic*. In that case wRw' is read as: w' is a deontically perfect alternative to w . Consequently, $w \models \Box A$ iff A is the case in all deontically perfect alternatives to w , and $w \models \Diamond A$ iff there is a deontically perfect alternative to w , in which A is true.

It is clear that in deontic logic $\Box A \rightarrow A$ and $A \rightarrow \Diamond A$ do not hold. This means that in general the accessibility relation R should not be reflexive. On the other hand, $\Box A \rightarrow \Diamond A$ should be valid in deontic logic. A necessary and sufficient condition on R in order to achieve this is that for each world w in a given Kripke model M there is a w' in M such that wRw' . This condition also rules out $\Box A \wedge \Box \neg A$ (something is obligatory and forbidden).

However, certain theorems are not dependent upon any condition concerning R . Some of these theorems have been viewed with suspicion because of their paradoxical appearance as deontic principles. For example, A. Ross illustrated the oddity of $\Box A \rightarrow \Box(A \vee B)$ by substituting 'I mail a letter' for A and 'I burn the letter' for B . The result 'if I ought to mail a letter, then I ought to mail or burn it' is known as *Ross's paradox*. A similar substitution may reveal the strangeness of $\Diamond A \rightarrow \Diamond(A \vee B)$. (See Exercise 6.6.) However, although $\Box(A \vee B)$ is true if $\Box A$ is true, according to Grice's [16] conversation rules, discussed in Section 2.10.2, it is simply misleading to say $\Box(A \vee B)$, when one knows $\Box A$. For more information on deontic logic the reader is referred to Hilpinen [18].

Also the \Box -axiom for $S4$, $\Box A \rightarrow \Box\Box A$, is not in all worlds of all Kripke models true. The following Kripke model $M = \langle \{w_0, w_1, w_2\}, R, \models \rangle$ with $w_0 R w_1$, $w_1 R w_2$, but not $w_0 R w_2$, is a counterexample:



$M, w_0 \models \Box P$, because $M, w_1 \models P$. But $M, w_0 \not\models \Box\Box P$, because $w_0 R w_1$ and $M, w_1 \not\models \Box P$, the latter because $M, w_2 \not\models P$.

It is easy to see that $\Box A \rightarrow \Box\Box A$ holds precisely in those Kripke models $M = \langle W, R, \models \rangle$ in which the accessibility relation R is *transitive*, i.e., for all w, w', w'' in M , if $w R w'$ and $w' R w''$, then $w R w''$.

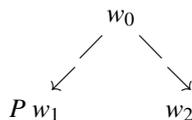
Proof. Let M be a Kripke model, w a world in M , and suppose $M, w \models \Box A$, i.e., for all w' in M , if $w R w'$, then $M, w' \models A$. (1)

We have to show that $M, w \models \Box\Box A$, i.e., for all w' in M , if $w R w'$, then $M, w' \models \Box A$. So, suppose that $w R w'$. (2)

We have to show that $M, w' \models \Box A$, i.e., for all w'' in M , if $w' R w''$, then $M, w'' \models A$. So, suppose that $w' R w''$. (3)

Assuming that R is transitive, it follows from (2) and (3) that $w R w''$. Hence, from (1): $M, w'' \models A$. \square

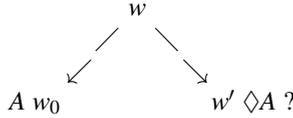
Finally, also the \Box -axiom for $S5$, $\Diamond A \rightarrow \Box\Diamond A$, is not in all worlds of all Kripke models true. The following Kripke model $M = \langle \{w_0, w_1, w_2\}, R, \models \rangle$ with $w_0 R w_1$ and $w_0 R w_2$, but not $w_2 R w_1$, is a counterexample:



$M, w_0 \models \Diamond P$, because $w_0 R w_1$ and $M, w_1 \models P$. But $M, w_0 \not\models \Box\Diamond P$, because $w_0 R w_2$ and $M, w_2 \not\models \Diamond P$.

It is not difficult to see that $\Diamond A \rightarrow \Box \Diamond A$ holds precisely in those Kripke models $M = \langle W, R, \models \rangle$ in which the accessibility relation R is transitive and *symmetric*, i.e., for all w and w' in M , if wRw' , then also $w'Rw$.

Proof. Let M be a Kripke model, w a world in M , and suppose $M, w \models \Diamond A$, i.e., there is some world w_0 in M such that wRw_0 and $M, w_0 \models A$. (1)



We have to show that $M, w \models \Box \Diamond A$, i.e., for all w' in M , if wRw' , then $M, w' \models \Diamond A$. So, suppose wRw' . (2)

We have to show that $M, w' \models \Diamond A$. Now, assuming that R is symmetric, it follows from (2) that also $w'Rw$. (3)

Assuming that R is transitive, it follows from (3) and (1) that $w'Rw_0$. And because $M, w_0 \models A$ (1), it follows that $M, w' \models \Diamond A$. \square

We collect the preceding results in the following theorem.

Theorem 6.2.

For every Kripke model $M = \langle W, R, \models \rangle$, $M \models \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$.

For every Kripke model $M = \langle W, R, \models \rangle$ with R reflexive, $M \models \Box A \rightarrow A$.

For every Kripke model $M = \langle W, R, \models \rangle$ with R transitive, $M \models \Box A \rightarrow \Box \Box A$.

For every Kripke model $M = \langle W, R, \models \rangle$ with R transitive and symmetric, $M \models \Diamond A \rightarrow \Box \Diamond A$.

Definition 6.10 (Kripke model for K^-). Let $M = \langle W, R, \models \rangle$ be a Kripke model. M is a Kripke model for KT iff R is reflexive. M is a Kripke model for $KT4 = S4$ iff R is reflexive and transitive. M is a Kripke model for $KT4E = S5$ iff R is reflexive, transitive and symmetric.

Definition 6.11 (Valid consequence in K^-). B is a *valid consequence* of premisses A_1, \dots, A_n in K^- := for all Kripke models M for K^- and for every world w in M , if $M, w \models A_1 \wedge \dots \wedge A_n$, then $M, w \models B$. **Notation:** $A_1, \dots, A_n \models B$ in K^- .

In case $n = 0$, i.e., there are no premisses, we say that B is *valid* in K^- , i.e., for all Kripke models M for K^- and for every world w in M , $M, w \models B$.

Notation: $\models B$ in K^- .

From Theorems 6.1 and 6.2 the following *soundness* theorem results, saying that any formula that may be logically deduced in K^- from given premisses is a valid consequence in K^- of those premisses:

Theorem 6.3 (Soundness of modal propositional logic).

If $A_1, \dots, A_n \vdash B$ in K^- , then $A_1, \dots, A_n \models B$ in K^- .

Proof. Suppose $A_1, \dots, A_n \vdash B$ in K^- , i.e., there is a finite schema of formulas with B as last one, such that every formula A in this schema is either one of A_1, \dots, A_n or an

axiom of classical propositional logic or the \Box -axiom of K or obtained by the rule Modus Ponens to two preceding formulas C and $C \rightarrow D$ in the schema or obtained by application of the rule for \Box to a preceding formula E in the schema such that $\vdash E$. We have to show that $A_1, \dots, A_n \models B$ in K . So, let M be a Kripke model, w be a world in M and suppose $M, w \models A_1 \wedge \dots \wedge A_n$. Notice that:

1. If A is an axiom of propositional logic or A is the \Box -axiom for K , then $M, w \models A$.
2. If $M, w \models C$ and $M, w \models C \rightarrow D$, then $M, w \models D$.
3. If $\vdash E$, then by Theorem 6.1 $\models \Box E$.

Hence, from 1, 2 and 3: $A_1, \dots, A_n \models B$ in K .

The proofs for KT , $S4$ and $S5$ are similar. □

Exercise 6.7. Prove that $A \not\models \Box A$, although by Theorem 6.1: if $\models A$, then $\models \Box A$.

Exercise 6.8. Prove or refute: a) $\models \Box(A \wedge B) \Leftrightarrow (\Box A \wedge \Box B)$;

b) $\models \Box(A \vee B) \Leftrightarrow (\Box A \vee \Box B)$; c) $\models \Diamond(A \vee B) \Leftrightarrow (\Diamond A \vee \Diamond B)$.

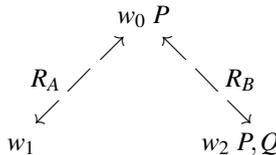
6.4 Epistemic logic

In *epistemic logic* $\Box A$ is read as ‘I know that A ’. More generally, $\Box_i A$ is read as ‘agent i knows that A ’, if one wants to consider more than one agent. Then $wR_i w'$ is read as: in world w agent i considers – on the ground of his knowledge – world w' as an (epistemic) alternative.

Because of the validity of $(\Box(A \rightarrow B) \wedge \Box A) \rightarrow \Box B$, epistemic logic is not concerned with actual occurrent knowledge, but with virtual or *implicit knowledge*. If a knower (or agent) knows A and $A \rightarrow B$, he or she also knows B , at least in principle, although one may not explicitly be aware of this.

In epistemic logic, one frequently uses K (Knowing) instead of the \Box -operator. For instance, K_A for ‘A(lice) knows A ’ and K_B for ‘B(ob) knows A ’.

As an example with two agents, consider the following state of affairs: A(lice) works in an office without windows, it is raining (P), but as far as Alice knows also $\neg P$ might be the case. B(ob) works in an office with windows, has been informed that it will rain all day and considers it possible that an important letter will arrive today (Q). We may model this state of affairs by the following Kripke model $M = \langle \{w_0, w_1, w_2\}, R_A, R_B, \models \rangle$ with $w_0 R_A w_1$ and $w_0 R_B w_2$, R_A and R_B both reflexive, transitive and symmetric, $w_0 \models P$, but $w_1 \not\models P$, $w_2 \models P$ and $w_2 \models Q$.



Clearly, $M, w_0 \not\models K_A P$ (in world w_0 of model M , Alice does not know P), because $w_0 R_A w_1$ and $w_1 \not\models P$ (Alice can imagine w_1 in which it does not rain). But $M, w_0 \models$

$K_B P$ (in world w_0 of model M , Bob knows P), because P holds in both worlds Bob can imagine: w_0 and w_2 .

$M, w_0 \models K_A(K_B P \vee K_B \neg P)$ (Alice knows in world w_0 of model M that Bob knows if P holds), because $M, w_0 \models K_B P$ (Bob knows in world w_0 that P) and $M, w_1 \models K_B \neg P$ (because from world w_1 Bob can only imagine w_1).

$M, w_0 \models \neg K_B(\neg K_A P)$ (Bob does not know in w_0 that Alice does not know P), because $w_0 R_B w_2$ and $M, w_2 \models K_A P$ (from world w_0 Bob can imagine world w_2 and in w_2 Alice knows P , because the only world she can imagine from w_2 is w_2 itself).

As this example suggests, epistemic logic can be used for the formal description of the knowledge of 'agents' in distributed systems. A nice illustration is the *muddy children puzzle*. See also Exercise 6.9.

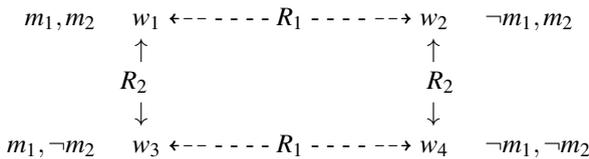
6.4.1 Muddy Children Puzzle; Reasoning about Knowledge

Imagine the following state of affairs. Two children are playing outside and their father asks them to come home. Both have mud on their foreheads, but they do not know themselves. Each child can see the other child, but not him- or herself; there are no mirrors. The father does not allow the children to talk to each other and says: at least one of you has mud on his forehead (P). If you *know* you have mud on your forehead, please step forward.

No child will step forward: each child sees the other child with mud on its forehead and considers it possible to be clean (without mud) himself. Notice that already before the statement of the father each child knows that P , but does not know that the other child knows P . After the statement of the father P has become common knowledge, in particular, now each child knows that the other child also knows P .

Since no child steps forward, the father repeats his request and asks again: if you *know* you have mud on your forehead, please step forward. Now both children step forward. Why? Because they can perfectly reason about knowledge: if there were only one child with mud, after the first statement/request of the father this child would know that he is the one with mud and step forward. Since no one stepped forward, there must be (at least) two children with mud.

We may model the state of affairs before the statement of the father by the following Kripke model M , where m_i stands for 'child i , $i = 1, 2$, has mud on his forehead', and R_i is the accessibility relation for child i .

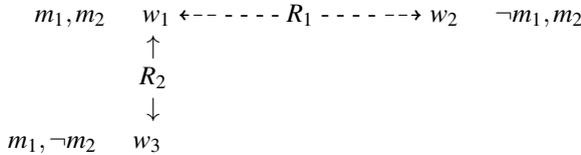


Before the statement of the father there are four possible worlds/situations, described by w_1, w_2, w_3 and w_4 . For instance, in world w_1 child 1 sees that child 2 has mud on his forehead, but child 1 can imagine to have no mud himself, i.e.,

world w_2 is accessible from world w_1 for child 1: $w_1 R_1 w_2$. Conversely, from world w_2 child 1 can imagine world w_1 : $w_2 R_1 w_1$. In a similar way, from world w_1 child 2 can easily imagine world w_3 and conversely: $w_1 R_2 w_3$ and $w_3 R_2 w_1$. The relations R_1 and R_2 are reflexive, transitive and symmetric.

Notice that $M, w_1 \models K_1 m_2 \wedge K_2 m_1$. In addition, for each world w in M , $M, w \models \neg K_1 m_1$ and $M, w \models \neg K_2 m_2$.

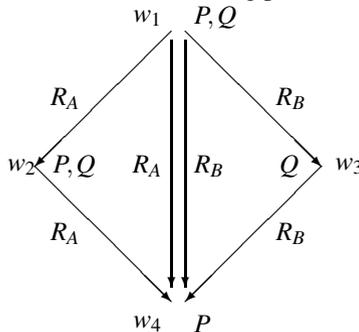
By the statement P of the father, world w_4 is eliminated and only three possible worlds are left, as described by the following Kripke model M' :



After the first statement P of the father, child 1 still does not know that he has mud on his forehead, because it sees child 2 with mud. This corresponds with $M', w_1 \models \neg K_1 m_1$. Similarly, $M, w_1 \models \neg K_2 m_2$.

If there would be only one child with mud, that is, if w_2 or w_3 would be the actual world, then, of course, after the first statement P of the father, the child with mud would *know* he has mud on his forehead, since he sees that the other child has no mud on his forehead. This corresponds with $M', w_2 \models K_2 m_2$ and $M', w_3 \models K_1 m_1$. So, if after the first statement/request of the father no child steps forward, each perfect logician will know that there must be at least two children with mud, in other words that world w_2 and w_3 do not occur and that only world w_1 is left. The new state of affairs is described by the Kripke model M'' containing only one possible world, i.e., w_1 . And $M'', w_1 \models K_1 m_1 \wedge K_2 m_2$.

Exercise 6.9 (J.J.Ch. Meyer). Consider the following Kripke model M consisting of four possible worlds w_1, w_2, w_3, w_4 , two agents A(lice) and B(ob) with reflexive and transitive accessibility relations R_A and R_B respectively, and suppose that R_A, R_B and \models are defined as indicated in the following picture.



Check that

$$\begin{aligned}
 M, w_1 \models Q & \quad , M, w_1 \models \neg K_A Q \quad , M, w_1 \models \neg K_B K_A P \quad , M, w_1 \models \neg K_B \neg K_A Q, \\
 M, w_1 \models K_A P & \quad , M, w_1 \models \neg K_B Q \quad , M, w_1 \models \neg K_A \neg K_B P \quad , M, w_1 \models \neg K_A \neg K_B Q, \\
 M, w_1 \models \neg K_B P & \quad , M, w_1 \models K_A K_A P \quad , M, w_1 \models K_A \neg K_A Q \quad , M, w_1 \models K_B \neg K_B Q.
 \end{aligned}$$

6.5 Tableaux for Modal Logics

A *tableaux system* for the modal logics K , KT and $S4$ is obtained by adding T and F rules for the modal operator \Box to the T and F rules for the connectives $\rightarrow, \wedge, \vee, \neg$ of classical propositional logic, given in Section 2.8 and listed below. Now TA is read as: A is true in world w ; and FA as: A is false in world w . We do not give the tableaux rules for $S5$ here, because they are complicated and hence somewhat artificial; the interested reader is referred to de Swart [36]. In the *tableaux rules* below, S is a *sequent*, i.e., a set of T - or F -signed formulas.

$$\begin{array}{ll}
 T\wedge & \frac{S, T B \wedge C}{S, TB, TC} \qquad F\wedge & \frac{S, F B \wedge C}{S, FB \mid S, FC} \\
 T\vee & \frac{S, T B \vee C}{S, TB \mid S, TC} \qquad F\vee & \frac{S, F B \vee C}{S, FB, FC} \\
 T\rightarrow & \frac{S, T B \rightarrow C}{S, FB \mid S, TC} \qquad F\rightarrow & \frac{S, F B \rightarrow C}{S, TB, FC} \\
 T\neg & \frac{S, T \neg B}{S, FB} \qquad F\neg & \frac{S, F \neg B}{S, TB}
 \end{array}$$

For K there is no $T\Box$ rule, but only a $F\Box$ rule: $F\Box \frac{S, F\Box A}{S_{\Box}, FA}$

For KT (or KM or M or T) the $T\Box$ and $F\Box$ rules are:

$$\begin{array}{ll}
 T\Box & \frac{S, T\Box A}{S, T\Box A, TA} \qquad F\Box & \frac{S, F\Box A}{S_{\Box}, FA}
 \end{array}$$

and for $S4$ these rules are:

$$\begin{array}{ll}
 T\Box & \frac{S, T\Box A}{S, T\Box A, TA} \qquad F\Box & \frac{S, F\Box A}{S_{T\Box}, FA}
 \end{array}$$

where $S_{\Box} := \{TB \mid T\Box B \in S\}$ and $S_{T\Box} := \{T\Box B \mid T\Box B \in S\}$, i.e., S_{\Box} contains all expressions TB for which $T\Box B$ occurs in S and $S_{T\Box}$ is the set of all expressions $T\Box B$ which occur in S . We have drawn a line in the rules $F\Box$ in order to stress that in the transition from S to S_{\Box} and $S_{T\Box}$, resp., some signed formulas may get lost.

The T - and F -rules for the propositional connectives follow the truth tables for these connectives. For instance, $B \rightarrow C$ is true in world w ($T B \rightarrow C$) iff B is false in w (FB) or C is true in w (TC); and $B \rightarrow C$ is false in w ($F B \rightarrow C$) iff B is true in w (TB) and C is false in w (FC). For obvious reasons the rules $T \rightarrow, T\vee$ and $F\wedge$ are called *split-rules*.

The intuitive motivation behind the T -rule for \Box is this one: if $\Box A$ is true in a world w , then also A will be true in world w , at least if w is accessible from itself, i.e., when R is reflexive. So, this $T\Box$ rule will apply in KT and in $S4$, but not in K .

The intuitive motivation behind the F -rule for \Box is the following one: if $\Box A$ is false in world w , then there must be a world w' , accessible from w , in which A is false. Since F -signed formulas (which are supposed to be false in w) do not have

to be false in w' , these formulas are not copied. In general, also T -signed formulas (which are supposed to be true in w) do not have to be true in w' and hence are not copied. There is one exception: If a T -signed formula $\Box B$ is true in w , then B will be true in w' ; and even $\Box B$ will be true in w' , if the accessibility relation R is transitive. So, we have different $F\Box$ rules for K and KT on the one hand, and for $S4$ on the other hand.

$A_1, \dots, A_n \vdash' B$ (B is *tableau-deducible* from A_1, \dots, A_n) in K , KT or $S4$, resp., is defined in a similar way as in Definition 2.18, the only difference being that there are two more rules for \Box .

Example 6.2. Let us verify that $\Box(A \rightarrow B) \vdash' \Box A \rightarrow \Box B$ in K . We construct a tableau starting with the premiss(es) T -signed and the putative conclusion F -signed; informally: we suppose the premisses are true and the putative conclusion false. Next we apply the T and F rules for the different connectives and modal operator.

$$\frac{\frac{\frac{T\Box(A \rightarrow B), F(\Box A \rightarrow \Box B)}{T\Box(A \rightarrow B), T\Box A, F\Box B}}{T(A \rightarrow B), TA, FB}}{FA, TA, FB \mid TB, TA, FB}$$

Since both ‘branches’ close, i.e., contain TC and FC for some formula C , this schema is by definition a tableau-deduction (in K) of $\Box A \rightarrow \Box B$ from $\Box(A \rightarrow B)$. Therefore, we have shown that $\Box(A \rightarrow B) \vdash' \Box A \rightarrow \Box B$ (in K), i.e., one can construct such a tableau-deduction. Informally: the supposition that the premisses are true and the conclusion false turns out to be untenable.

Example 6.3. Let us verify that $\vdash' \Box A \rightarrow A$ in KT , but not in K : $F(\Box A \rightarrow A)$
 $T\Box A, FA$
 TA, FA

The only ‘branch’ is closed, and hence $\vdash' \Box A \rightarrow A$ in KT .

Notice that this tableau-proof does not hold in K , because there is no $T\Box$ rule for K . If we make a tableau in K for $\Box A \rightarrow A$ we find:

$$\frac{F(\Box A \rightarrow A)}{T\Box A, FA} \quad w$$

which does not close. In fact, we have constructed a Kripke counterexample $M = \langle \{w\}, R, \models \rangle$ in K , with, by definition, not wRw and $w \not\models A$, corresponding with the occurrence of FA in w . $M, w \models \Box A$, since there is no world accessible from w in which A is not true. But $M, w \not\models A$.

Example 6.4. Let us verify that $\vdash' \Box A \rightarrow \Box\Box A$ in $S4$, but not in KT :

$$\frac{\frac{\frac{F(\Box A \rightarrow \Box\Box A)}{T\Box A, F\Box\Box A}}{T\Box A, TA, F\Box\Box A}}{T\Box A, F\Box A} \quad \frac{T\Box A, TA, F\Box A}{T\Box A, FA}}{T\Box A, TA, FA}$$

The only ‘branch’ of this tableau is closed, and hence $\vdash' \Box A \rightarrow \Box\Box A$ in $S4$.

Notice that this tableau-proof does not hold in KT . A tableau starting with $F(\Box A \rightarrow \Box\Box A)$ in KT will look as follows and does not close:

$F(\Box A \rightarrow \Box\Box A)$	
$T\Box A, F\Box\Box A$	$w_0 A$
$T\Box A, TA, F\Box\Box A$	\downarrow
<u>$TA, F\Box A$</u>	$w_1 A$
	\downarrow
FA	w_2

In fact, we have constructed a Kripke countermodel $M = \langle \{w_0, w_1, w_2\}, R, \models \rangle$ in KT , with $w_0 R w_1, w_1 R w_2$, but not $w_0 R w_2$, R reflexive, but not transitive, and by definition $w_0 \models A, w_1 \models A$, but $w_2 \not\models A$, corresponding with the occurrence of TA in w_0 and w_1 and the occurrence of FA in w_2 . Then, corresponding with the occurrence of $T\Box A$ in w_0 , $M, w_0 \models \Box A$, since $M, w_0 \models A$ and $M, w_1 \models A$, but, corresponding with the occurrence of $F\Box\Box A$ in w_0 , $M, w_0 \not\models \Box\Box A$, since $M, w_1 \not\models \Box A$. Notice that if R were transitive, we would not have that $M, w_0 \models \Box A$.

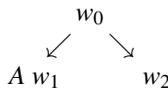
Example 6.5. We shall try to construct a tableau proof of the $S5$ -axiom $\Diamond A \rightarrow \Box\Diamond A$ in $S4$. So, we start with $F(\Diamond A \rightarrow \Box\Diamond A)$:

$F(\Diamond A \rightarrow \Box\Diamond A)$
$T\Diamond A, F\Box\Diamond A$
$T\neg\Box\neg A, F\Box\neg\Box\neg A$
$F\Box\neg A, F\Box\neg\Box\neg A$

At this point there are two possibilities to continue: we may proceed with $F\Box\neg A$ losing the second F -signed formula, or we may proceed with $F\Box\neg\Box\neg A$ losing the first F -signed formula. Either way, we do not get closure and hence we do not find a tableau proof in $S4$ of $\Diamond A \rightarrow \Box\Diamond A$:

$F\neg A$	$F\neg\Box\neg A$
TA	$T\Box\neg A$
	$T\Box\neg A, T\neg A$
	$T\Box\neg A, FA$

We shall call the resulting tree the *search tree* for the conjecture $\vdash' \Diamond A \rightarrow \Box\Diamond A$ in $S4$. From this search tree one can immediately read off a Kripke counterexample $M = \langle \{w_0, w_1, w_2\}, R, \models \rangle$ in $S4$ for this formula, with, by definition, $w_0 R w_1, w_0 R w_2$, R reflexive and transitive, but not symmetric, and $w_1 \models A$, corresponding with the occurrence of TA in w_1 :



Then, corresponding with the occurrence of $T\Diamond A$ in w_0 , $M, w_0 \models \Diamond A$, since $w_0 R w_1$ and $M, w_1 \models A$. But, corresponding with the occurrence of $F\Box\Diamond A$ in w_0 , $M, w_0 \not\models \Box\Diamond A$

$\Box\Diamond A$, since w_0Rw_2 and $M, w_2 \not\models \Diamond A$, corresponding with the occurrence of $F\Diamond A$ in w_2 . Notice: if R were symmetric, we would have $M, w_2 \models \Diamond A$, because in that case symmetry would guarantee w_2Rw_0 and next transitivity would guarantee w_2Rw_1 .

Example 6.6. The following tableau \mathcal{T} with initial branch $\mathcal{B}_0 = \{T\Box(P \wedge Q), F(\Box P \wedge (\Box Q \vee \Box R))\}$ is a tableau-deduction of $\Box P \wedge (\Box Q \vee \Box R)$ from $\Box(P \wedge Q)$ in K :

$$\begin{array}{c} T\Box(P \wedge Q), F(\Box P \wedge (\Box Q \vee \Box R)) \\ T\Box(P \wedge Q), F\Box P \mid T\Box(P \wedge Q), F(\Box Q \vee \Box R) \\ \hline T\Box(P \wedge Q), F\Box P \mid T\Box(P \wedge Q), F\Box Q, F\Box R \\ \hline T(P \wedge Q), FP \mid T(P \wedge Q), FQ \\ TP, TQ, FP \mid TP, TQ, FQ \end{array}$$

Notice that both branches are closed, i.e., contain for some formula C both TC and FC . Also notice that in the right branch, instead of applying the $F\Box$ rule to $F\Box Q$, we might also have applied the $F\Box$ rule to $F\Box R$, in which case the right branch would finish with TP, TQ, FR and hence would not close.

Let branch $\mathcal{B}_1 = \mathcal{B}_0 \cup \{F\Box P\}$ and branch $\mathcal{B}_2 = \mathcal{B}_0 \cup \{F(\Box Q \vee \Box R)\}$. Then tableau $\mathcal{T}_1 = \{\mathcal{B}_1, \mathcal{B}_2\}$ is called a *one-step expansion* in K of tableau $\mathcal{T}_0 = \{\mathcal{B}_0\}$.

Let branch $\mathcal{B}_{11} = \mathcal{B}_1$ and branch $\mathcal{B}_{21} = \mathcal{B}_2 \cup \{F\Box Q, F\Box R\}$. Then tableau $\mathcal{T}_2 = \{\mathcal{B}_{11}, \mathcal{B}_{21}\}$ is called a *one-step expansion* in K of tableau \mathcal{T}_1 .

Let branch $\mathcal{B}_{111} = \mathcal{B}_{11}^* \cup \{T(P \wedge Q), FP\}$ and let $\mathcal{B}_{211} = \mathcal{B}_{21}^* \cup \{T(P \wedge Q), FQ\}$, where \mathcal{B}^* indicates that the formulas in \mathcal{B} do not count towards closure anymore. Then tableau $\mathcal{T}_3 = \{\mathcal{B}_{111}, \mathcal{B}_{211}\}$ is called a *one-step expansion* in K of \mathcal{T}_2 .

Finally, let branch $\mathcal{B}_{1111} = \mathcal{B}_{111} \cup \{TP, TQ\}$ and $\mathcal{B}_{2111} = \mathcal{B}_{211} \cup \{TP, TQ\}$. Then tableau $\mathcal{T}_4 = \{\mathcal{B}_{1111}, \mathcal{B}_{2111}\}$ is called a *one-step expansion* in K of \mathcal{T}_3 .

Definition 6.12 ((Tableau) Branch). (a) A *tableau branch* is a set of signed formulas. A branch is *closed* if it contains signed formulas TA and FA for some formula A . A branch that is not closed is called *open*.

(b) Let \mathcal{B} be a branch and TA , resp. FA , a signed formula occurring in \mathcal{B} . TA , resp. FA , is *fulfilled* in \mathcal{B} if (i) A is atomic, or (ii) \mathcal{B} contains the bottom formulas in the application of the corresponding T or F rule to A , and in case of the rules $T\vee$, $F\wedge$ and $T\rightarrow$, \mathcal{B} contains one of the bottom formulas in the application of these rules.

(c) A branch \mathcal{B} is *completed* if \mathcal{B} is closed or every signed formula in \mathcal{B} is fulfilled in \mathcal{B} .

Definition 6.13 (Tableau). (a) A set \mathcal{T} of branches is a *tableau* in $K-$ with initial branch \mathcal{B}_0 if there is a sequence $\mathcal{T}_0, \mathcal{T}_1, \dots, \mathcal{T}_n$ such that $\mathcal{T}_0 = \{\mathcal{B}_0\}$, each \mathcal{T}_{i+1} is a one-step expansion in $K-$ of \mathcal{T}_i ($0 \leq i < n$) and $\mathcal{T} = \mathcal{T}_n$.

(b) We say that a finite \mathcal{B} has tableau \mathcal{T} if \mathcal{T} is a tableau with initial branch \mathcal{B} .

(c) A tableau \mathcal{T} in $K-$ is *open* if some branch \mathcal{B} in it is open, otherwise \mathcal{T} is *closed*.

(d) A tableau is *completed* if each of its branches is completed; informally, no application of a tableau rule can change the tableau.

Definition 6.14 (Tableau-deduction; Tableau-proof).

(a) A *tableau-deduction* of B from A_1, \dots, A_n in K^- is a tableau \mathcal{T} in K^- with $\mathcal{B}_0 = \{TA_1, \dots, TA_n, FB\}$ as initial branch, such that all branches of \mathcal{T} are closed.

In case $n = 0$, i.e., there are no premisses A_1, \dots, A_n , this definition reduces to:

(b) A *tableau-proof* of B in K^- is a tableau \mathcal{T} in K^- with $\mathcal{B}_0 = \{FB\}$ as initial sequent, such that all branches of \mathcal{T} are closed.

Definition 6.15 (Tableau-deducible; Tableau-provable).

(a) B is *tableau-deducible* from A_1, \dots, A_n in K^- := there exists a tableau-deduction of B from A_1, \dots, A_n in K^- . **Notation:** $A_1, \dots, A_n \vdash' B$ in K^- .

(b) B is *tableau-provable* in K^- := there exists a tableau-proof of B in K^- .

Notation: $\vdash' B$ in K^- . And for Γ a (possibly infinite) set of formulas,

(c) B is *tableau-deducible* from Γ in K^- := there exists a finite list A_1, \dots, A_n of formulas in Γ such that $A_1, \dots, A_n \vdash' B$ in K^- . **Notation:** $\Gamma \vdash' B$ in K^- .

Example 6.7. a) As seen in Example 6.2, $\Box(A \rightarrow B) \vdash' (\Box A \rightarrow \Box B)$ in K .

b) As seen in Example 6.3, $\Box A \vdash' A$ in KT , or, equivalently, $\vdash' \Box A \rightarrow A$ in KT .

c) As seen in Example 6.4, $\Box A \vdash' \Box \Box A$ in $S4$ or, equivalently, $\vdash' \Box A \rightarrow \Box \Box A$ in $S4$.

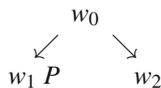
Example 6.8. We wonder whether $\Diamond \Box P \vdash' \Box \Diamond P$ in $S4$. We start a tableau with $T \Diamond \Box P, F \Box \Diamond P$ in $S4$:

$$\begin{array}{c} T \Diamond \Box P, F \Box \Diamond P \\ F \Box \neg \Box P, F \Box \Diamond P \end{array}$$

We may continue with $F \Box \neg \Box P$, losing $F \Box \Diamond P$ and we may continue with $F \Box \Diamond P$, losing $F \Box \neg \Box P$. If one of these two options would give closure, we would have found a tableau deduction of $\Box \Diamond P$ from $\Diamond \Box P$ in $S4$. However, it turns out that either way does not give closure:

$$\begin{array}{ccc} & T \Diamond \Box P, F \Box \Diamond P & \\ & F \Box \neg \Box P, F \Box \Diamond P & \\ & \swarrow \quad \searrow & \\ F \neg \Box P & & F \Diamond P \\ T \Box P & & T \Box \neg P \\ T \Box P, TP & & T \Box \neg P, T \neg P \\ & & T \Box \neg P, FP \end{array}$$

We shall call the resulting tree the *search tree* for the conjecture $\Diamond \Box P \vdash' \Box \Diamond P$ in $S4$. From this search tree with both branches open we may immediately read off a Kripke counterexample $M = \langle \{w_0, w_1, w_2\}, R, \models \rangle$ in $S4$ with, by definition, $w_0 R w_1$, $w_0 R w_2$, R reflexive and transitive, $w_1 \models P$, corresponding with the occurrence of TP in w_1 , and $w_2 \not\models P$, corresponding with the occurrence of FP in w_2 :



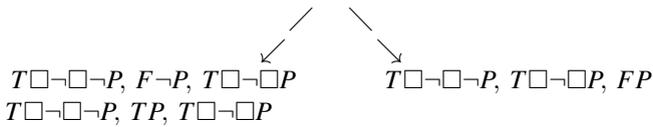
Clearly, $M, w_0 \models \Diamond \Box P$, since $M, w_1 \models \Box P$, but $M, w_0 \not\models \Box \Diamond P$, since $M, w_2 \not\models \Diamond P$.

Example 6.9. We wonder whether $\Box\Diamond P \vdash' \Diamond\Box P$ in $S4$.

We start a tableau with $T\Box\Diamond P, F\Diamond\Box P$ in $S4$, i.e.,

$$\begin{aligned} & T\Box\neg\Box\neg P, F\neg\Box\neg\Box P \\ & T\Box\neg\Box\neg P, T\neg\Box\neg P, F\neg\Box\neg\Box P \\ & T\Box\neg\Box\neg P, T\neg\Box\neg P, T\Box\neg\Box P \\ & T\Box\neg\Box\neg P, F\Box\neg P, T\Box\neg\Box P \\ & T\Box\neg\Box\neg P, F\Box\neg P, T\Box\neg\Box P, T\neg\Box P \\ & T\Box\neg\Box\neg P, F\Box\neg P, T\Box\neg\Box P, F\Box P \quad (*) \end{aligned}$$

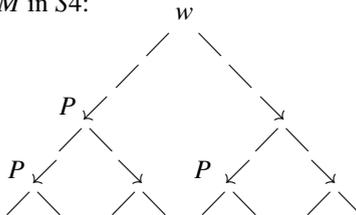
At this stage we have applied the $T\Box$ rule as many times as possible and we now have two signed formulas of the form $F\Box$. If we apply the $F\Box$ rule to either one of them, we loose the other. So, there are two possibilities to go on; if one of them would give closure, we would have a tableau deduction of $\Diamond\Box P$ from $\Box\Diamond P$.



$T\Box\neg\Box\neg P$ will give $T\neg\Box\neg P$ and next $F\Box\neg P$ again, and $T\Box\neg\Box P$ will give $T\neg\Box P$ and next $F\Box P$. So, the tableau will continue with

$$F\Box\neg P, TP, F\Box P \qquad F\Box\neg P, F\Box P, FP.$$

So, we are essentially back at line (*) with $F\Box\neg P$ and $F\Box P$, from where the situation repeats itself. However, no branch will ever close and we read off the following Kripke counterexample M in $S4$:



Clearly, $M, w \models \Box\Diamond P$, i.e., for every w' in M with wRw' there is a w'' in M such that $w'Rw''$ and $M, w'' \models P$; but $M, w \not\models \Diamond\Box P$, i.e., there is no w' in M with wRw' such that for all all w'' in M , if $w'Rw''$, then $M, w'' \models P$. Hence, $\Box\Diamond P \not\models \Diamond\Box P$.

The examples given above suggest a general procedure which, given a conjecture $A_1, \dots, A_n \vdash' B$ in $K-$, will either construct a tableau-deduction of B from the premisses A_1, \dots, A_n in $K-$ or yield a Kripke counterexample in $K-$. We shall describe this procedure in more detail in Section 6.7 and prove that the three notions $A_1, \dots, A_n \vdash B$ in $K-$, $A_1, \dots, A_n \models B$ in $K-$, and $A_1, \dots, A_n \vdash' B$ in $K-$, are equivalent.

Exercise 6.10. Translate the following argument in the language of modal propositional logic and either construct a tableau-deduction in K of the putative conclusion from the premisses or construct a Kripke counterexample in K .

It is not the case that: if John works hard $[W]$, then he will necessarily succeed $[S]$. Therefore, it is possible that: if John works hard, then he will not succeed.

Exercise 6.11. Translate the following argument in the language of modal propositional logic and either construct a tableau-deduction in K of the putative conclusion from the premisses or construct a Kripke counterexample in K .

It is possible that: if John fails [J], then he will give a party [P].

Therefore, if John fails, then it is possible that he will give a party.

Exercise 6.12. Prove that i) $K_i(A \vee B), K_i \neg A \vdash K_i B$ in K ,

but ii) $A \vee B, K_i \neg A \not\vdash K_i B$ in K , neither in $S4$.

This explains the paradox in Exercise 2.70: let A stand for 'the prisoner will be hanged on Monday, Tuesday, Wednesday or Thursday' and let B stand for 'the prisoner will be hanged on Friday'. Then $A \vee B$ is the judge's statement that the prisoner would hang one day this week. Read $K_i E$ as 'prisoner i knows (on Friday morning) that E '. See also the answer to Exercise 2.70.

Exercise 6.13. Prove or refute in K : a) $\Box A \vee \neg \Box A$; b) $\Box A \vee \Box \neg A$.

Prove or refute in KT : c) $\Diamond A \vee \neg \Diamond A$; d) $\Diamond A \vee \Diamond \neg A$.

Exercise 6.14. Prove: $\vdash \Box A \rightarrow \Box(A \vee B)$ in K and $\vdash \Diamond A \rightarrow \Diamond(A \vee B)$ in K (cf. Exercise 6.6).

Exercise 6.15. Prove that K, KT and $S4$ have the *disjunction property*:

if $\vdash \Box A \vee \Box B$, then $\vdash \Box A$ or $\vdash \Box B$.

Exercise 6.16. Prove or refute in KT : a) $\Diamond P \rightarrow \Diamond \Diamond P$; b) $\Diamond \Diamond P \rightarrow \Diamond P$.

Prove or refute in $S4$: c) $P \rightarrow \Box \Diamond P$; d) $(P \rightarrow Q) \rightarrow \neg \Diamond (P \wedge \neg Q)$.

6.6 Applications of Possible World Semantics

6.6.1 Direct Reference

There are at least two problems in the traditional theory of meaning:

1. In the traditional view, a proper name, like 'Jane', is identified with a description, such as 'the woman John is married to'. Now suppose that John is a bachelor. Then it would follow that Jane does not exist. This example makes clear that a person can be referred to by his or her name even if the description of the person in question does not apply to that person.
2. According to the traditional theory, a tiger, for instance, is identified with an object which has certain properties, among which the property of having sharp teeth. Consequently, the statement 'tigers have sharp teeth' is analytic; this seems to be counter-intuitive.

In the traditional theory, the conjunction of properties which a tiger is supposed to have is called the intension of the word 'tiger' and is supposed to be the *essence* of tiger. In the traditional theory as well, intension determines extension. Similarly, in the traditional view, the proper name 'Aristotle' is identified with a description

such as ‘the most well-known man who studied under Plato’. As a consequence, the proposition ‘Aristotle studied under Plato’ would be an analytic truth. This is again against our intuition.

Typical of the theory of direct reference is the position, held by Kripke, Donnellan and others, that proper names and nouns standing for natural kinds refer independently of identifying descriptions. In his paper [9], Donnellan distinguished between two kinds of use for definite descriptions – the *attributive use* and the *referential use*. In order to make this distinction clear, Donnellan considered the use of the definite description ‘Smith’s Murderer’ in the following two cases.

Suppose first that we come upon poor Smith foully murdered. From the brutal manner of the killing and the fact that Smith was the most lovable person in the world, we might exclaim ‘Smith’s murderer is insane’. I will assume, to make it a simpler case, that in a quite ordinary sense we do not know who murdered Smith. . . . This, I shall say, is an *attributive use* of the definite description. [[9], 285-286]

So, in the case of the attributive use, the speaker wants to say something about whoever or whatever fits the description even if he does not know who or what that is. On the other hand,

Suppose that Jones has been charged with Smith’s murder and has been placed on trial. Imagine that there is a discussion of Jones’ odd behavior at his trial. We might sum up our impression of his behavior by saying ‘Smith’s murderer is insane’. If someone asks to whom we are referring by using this description, the answer here is ‘Jones’. This, I shall say, is a *referential use* of the definite description.

So, if the description ‘Smith’s murderer’ is used referentially, the speaker is referring to Jones, even in the case that Jones turns out to be innocent. Note that in this case the description refers to Jones although it does not apply to Jones. To give another example, suppose someone asks me at a party who Mr. X is. I answer ‘the man at the door with a glass of sherry in his hand’. Now suppose that the person referred to actually has a glass of white wine in his hand. Again the description may refer successfully without applying to the object referred to. These examples make clear that descriptions, when used referentially, do not always apply to the object they refer to. When using a description referentially, we have a definite object in mind whether or not it does fit the description.

According to the theory of *direct reference*, brought out by Keith Donnellan, Saul Kripke and others, proper names, like ‘Aristotle’, ‘Thales’ and ‘Jane’, and nouns standing for natural kinds, like ‘gold’, ‘water’ and ‘tiger’, have no intension (Sinn) in the traditional sense, but only have reference; and this reference is established by a *causal chain* rather than by an associated description. For example, the reference to the person called ‘Aristotle’ is determined by a causal chain as follows. The person in question is given a name in a ‘baptism’ with the referent present. Next this name is handed on from speaker to speaker. It is in this way that we use the name ‘Aristotle’ referring to the person in question. We do not have to have any description of Aristotle; the information ‘Aristotle was a philosopher’ may be completely new to the one who is using the name ‘Aristotle’.

It is typical of the theory of direct reference that proper names, like 'Jane', refer to some definite object, even when the description we supply, such as 'the woman John is married to', does not apply to that object. This description may help us fix the reference, but it should not be taken to be the meaning of the name. And a similar view is held for nouns standing for natural kinds, like 'gold', 'water' and 'tiger'. The meaning of the word 'tiger' is its reference; identifying descriptions, such as 'a tawny-coloured animal with sharp teeth', only help us to fix the reference of this term.

Summarizing, according to the theory of direct reference, the meaning of a proper name or a natural kind term is its reference; the descriptions given in connection with these terms only help the hearer to pick out what the speaker has in mind.

6.6.2 Rigid Designators

In his paper *Naming and Necessity*, Kripke [22] in addition holds the view that a proper name, like 'Aristotle', is a *rigid designator*, i.e., it designates the very same object in all possible worlds in which this object exists. Thus, in the sentence 'Aristotle might have been a carpenter', the proper name 'Aristotle' refers to the same individual referred to in the sentence 'Aristotle was the philosopher who was a pupil of Plato and taught Alexander'. The definite description 'the most well-known man who studied under Plato', though it designates Aristotle in the actual world, may designate other individuals in other possible worlds; for it is possible that Aristotle did not study under Plato. Contrary to the traditional theory of meaning, according to the theory of direct reference, the statement 'Aristotle studied under Plato' is not necessarily true (and hence not analytic).

Now, if a and b are rigid designators and $a = b$ is true (in this world), then $a = b$ must be true in all worlds (accessible from this one) and hence $\Box(a = b)$ is true. So, it follows from the thesis that proper names are rigid designators that all true identity statements of the form $a = b$, where a and b are proper names, are necessarily true. In particular, it follows that 'Hesperus is Phosphorus (the morning star is the evening star)' and 'Tully is Cicero', if true (in this world) are necessarily true. On the other hand, we do not know a priori that Hesperus (the Morning Star) is Phosphorus (the Evening Star); this was discovered by empirical observation. Therefore, Kripke [23] claims in his paper *Identity and Necessity* that sentences like 'Hesperus is Phosphorus' and 'Tully is Cicero' if true (in this world) are *necessarily true* and at the same time are *a posteriori*.

Kripke extends his insights about proper names to nouns standing for natural kinds, such as 'gold', 'water' and 'tiger'. These nouns are rigid designators too, i.e., they refer to the same substance in all possible worlds in which this substance exists. Let us consider some interesting consequences of this point of view. 'Gold' being a rigid designator, the sentence 'gold is the element with atomic number 79', if true (in this world), will be true in all worlds (accessible from this one) and hence be necessarily true. Similarly, 'water' being a rigid designator, the sentence 'water

has the chemical structure H_2O ', if true (in this world), will be true in any world (accessible from this one) and hence be necessarily true. So both propositions, if true (in this world), are *necessarily true* and at the same time *a posteriori*. Kripke defines a sentence A to be *analytic* if it is both necessary and a priori. Consequently, sentences like 'Hesperus is Phosphorus', 'Tully is Cicero', 'gold is the element with atomic number 79' and 'water is H_2O ' are NOT analytic, since they are a posteriori, although necessarily true, if true (in this world).

Let stick S denote the standard meter in Paris. Then, by definition, stick S is one meter long. Therefore, the epistemological status of the statement 'stick S is one meter long' is that this statement is an *a priori* truth. Conceiving 'one meter' as a rigid designator, indicating the same length in all possible circumstances (worlds), the metaphysical status of 'stick S is one meter long' will be that of a *contingent* statement, since the length of stick S can vary with the temperature, humidity and so on. So, assuming that 'one meter' is a rigid designator, the sentence 'stick S is one meter long' is both *a priori* and *contingent*, i.e., not necessarily true.

Similarly, the sentence 'water boils at 100 degrees Celcius' will be a priori and at the same time contingent, i.e., not necessarily true, if we conceive '100 degrees Celcius' as a rigid designator.

6.6.3 *De dicto - de re distinction*

If one wants to translate the sentence

It is possible that a Republican will win

into a logical formula, it becomes evident that this sentence is ambiguous. Using \diamond for 'it is possible that', the predicate symbol R for 'being a Republican' and the symbol W for 'will win', there are two different translations of the sentence in question:

- (1) $\exists x[R(x) \wedge \diamond W(x)]$, and
- (2) $\diamond \exists x[R(x) \wedge W(x)]$.

(1) says, literally, that there is some particular individual who actually is a Republican and who may possibly win.

(2) says, literally, that it is possible that some Republican or other will win.

(1) is called the *de re* or referential reading of the sentence above. Typical of the *de re* reading is that the possibility operator \diamond occurs within the scope of the (existential) quantifier.

(2) is called the *de dicto* or non-referential reading of the sentence above. Typical of the *de dicto* reading is that the (existential) quantifier occurs within the scope of the possibility operator \diamond .

The example above demonstrates that sentences containing modalities such as 'possibly', 'necessarily', 'John believes that ...', etc., in combination with existential or universal quantifiers may give rise to ambiguities. Speaking in terms of possible worlds:

- (1) says that in the given world there is a person who is a Republican (in the given world) and who will win in some world accessible from the given one;
 (2) says that there is a world accessible from the given one in which there is a person who in that world is Republican and will win.

The proposition ‘John finds a unicorn’ can be properly translated as $\exists x[U(x) \wedge F(j,x)]$, where $U(a)$ stands for ‘ a is a unicorn’, j stands for ‘John’ and $F(a,b)$ stands for ‘ a finds b ’. But $\exists x[U(x) \wedge S(j,x)]$, where $S(a,b)$ stands for ‘ a seeks b ’ would be an improper translation of ‘John seeks a unicorn’, because the use of the existential quantifier commits us to an ontology in which unicorns do exist. Note that ‘John finds a unicorn’ and ‘John seeks a unicorn’ provide an extensional and an intensional context respectively (see Section 6.11).

In his paper [30], R. Montague develops a ‘categorical’ language in which ‘John seeks a unicorn’ can be properly translated.

6.6.4 Reasoning about Knowledge

Suppose three children, A(d), B(ob) and C(od), have played outside and two of them, say A and B, have mud on their forehead; they can see each other, but not themselves (there are no mirrors) and they do not communicate with each other. However, they are all perfect logicians! Let P be the proposition:

P : there is at least one child with mud on its forehead.

Notice that each child knows P , because A sees B, B sees A and C sees both A and B. But A does not know that B knows that P , because if A has no mud on its forehead, B sees nobody with mud. So, P is not *common knowledge*.

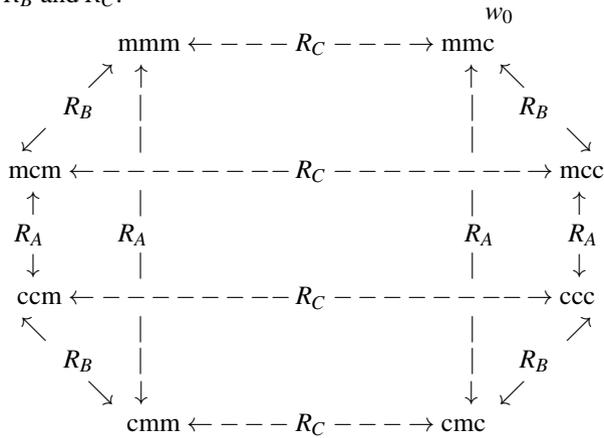
Now the father of the children announces P . By this announcement, P becomes *common knowledge*, in particular, everybody now knows that everybody knows P . For instance, A now knows that B knows P .

Next, the father asks each child (for the first time) to step forward if he *knows* to have mud on his forehead. What will happen? No child will step forward: A sees B with mud, B sees A with mud, and C sees both A and B with mud. So, no child has a reason to step forward.

Because after the first request no child steps forward, it becomes common knowledge that there must be at least two children with mud; if there were only one child with mud, this child would see no one else with mud and hence know he must be the one with mud. Consequently, if the father asks each child for the second time to step forward if he or she *knows* to have mud on the forehead, child A and B will step forward: A knows that there are at least two children with mud and only sees B with mud, and similarly for B.

Let m_A be the proposition ‘A has mud on his forehead’ and c_B the proposition ‘B is clean’. By definition, $w_{m_A c_B m_C}$, abbreviated by w_{mcm} or even mcm , is the world in which m_A , c_B and m_C are true, i.e., $w_{mcm} \models m_A \wedge \neg m_B \wedge m_C$. We may model the

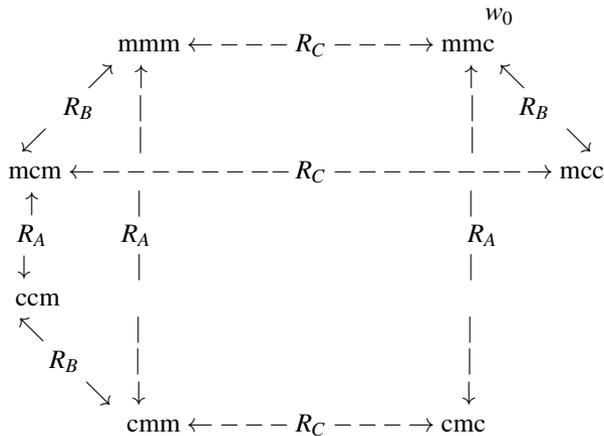
initial situation described above - before the father has said anything - by a Kripke model $M = \langle W, R_A, R_B, R_C, \models \rangle$ with eight possible worlds and three accessibility relations R_A, R_B and R_C .



In our story, the actual world is $w_0 = w_{mmc}$. Because the children cannot see themselves, A, for instance, cannot distinguish between w_{mmm} and w_{cmm} . So, the accessibility relations R_A, R_B and R_C are reflexive and symmetric.

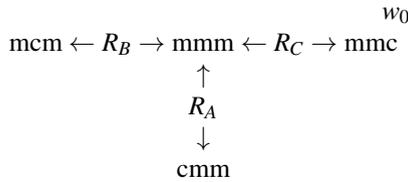
Notice that in world w_0 of this Kripke model M , A does not know that m_A , since A cannot distinguish between w_0 and w_{cmc} , in which m_A does not hold. In other words, $M, w_0 \not\models K_A m_A$, since $w_0 R_A w_{cmc}$, and $M, w_{cmc} \not\models m_A$. The proposition P , expressing that there is at least one child with mud, can now be rendered by $P = m_A \vee m_B \vee m_C$. In world w_0 of this Kripke model M , A does not know that B knows that P , because A cannot distinguish between w_0 and w_{cmc} , in which B does not know P , because B cannot distinguish between w_{cmc} and w_{ccc} . In other words, $M, w_0 \not\models K_A(K_B P)$, because $M, w_{cmc} \not\models K_B P$.

Once the father has announced the proposition P , each child eliminates the world w_{ccc} ; the new situation is now modelled by the Kripke model M' :



Notice: $M', w_0 \models K_A(K_B P)$, because $M', w_0 \models K_B P$ (B sees in w_0 that A has mud) and $M', w_{cmc} \models K_B P$ (B sees in w_{cmc} that A and C are clean).

In case that exactly one child, say A, has mud on his forehead, i.e., in world w_{mcc} of Kripke model M' , we have $M', w_{mcc} \models K_A m_A$, because the only world accessible for A from w_{mcc} is w_{mcc} , in which m_A is true (A sees that B and C are clean). Similarly, $M', w_{cmc} \models K_B m_B$ and $M', w_{ccm} \models K_C m_C$. So, after announcing the proposition P , if there were only one child with mud, the child in question would know that he has mud on his forehead and would step forward. Once it becomes clear that no child knows that he has mud on his forehead, it follows that the three possible worlds w_{mcc} , w_{cmc} and w_{ccm} are cancelled and the only remaining possible worlds are depicted in the following Kripke model M'' :



Now, clearly, $M'', w_0 \models K_A m_A \wedge K_B m_B$, so A and B will step forward. Similarly, $M'', w_{mcm} \models K_A m_A \wedge K_C m_C$ and $M'', w_{cmm} \models K_B m_B \wedge K_C m_C$.

If no child would step forward after the second request of the father, it would follow that the worlds w_{mmc} , w_{mcm} and w_{cmm} are eliminated from model M'' and only world w_{mmm} would remain, resulting in the Kripke model M''' , consisting of only one world w_{mmm} . And $M''', w_{mmm} \models K_A m_A \wedge K_B m_B \wedge K_C m_C$.

More generally, one may prove (see, for instance, Fagin, e.a. [10]):

Theorem 6.4. *If there are k , $k = 1, 2, \dots$, children with mud on the forehead, after announcing the proposition that there is at least one child with mud, the father has to state his request - to step forward once one knows that one has mud on the forehead - k times, before each child with mud knows that he has mud on his forehead. After i ($i < k$) rounds of questioning, it is common knowledge that at least $i + 1$ children have mud on their foreheads.*

6.6.5 Common Knowledge

As seen in Subsection 6.6.4 common knowledge plays an important role in the muddy children puzzle. But common knowledge is also relevant for reaching agreement or for coordinating actions. We shall illustrate this by the *coordinated attack problem* informally as follows:

There are two hills with a valley in between. On the hills are two divisions of an army, each with its own general and in the valley is the enemy. If both divisions attack the enemy simultaneously they will surely win, but if only one division attacks, it will be defeated and have serious losses. So each general wants to be absolutely

sure that both divisions attack at the same time. Say, general 1 wants to coordinate a simultaneous attack at dawn the next day and the generals are only able to communicate by means of a messenger (telephones are not available). The messenger, however, may get lost or may be captured by the enemy. How long will it take the generals to coordinate an attack?

Suppose general 1 sends a messenger with the message P (*we attack at dawn tomorrow morning*) to general 2. Initially, we have that K_1P and $\neg K_2P$, where K_i is the knowledge operator for general $i \in \{1, 2\}$. Even if the message is in fact delivered, general 1 does not know that it was delivered: $\neg K_1(K_2P)$; hence he cannot be sure that general 2 will attack simultaneously. So, given his state of knowledge, general 1 will not attack. General 2 knows this and does not want to take the risk of attacking alone; hence, he cannot attack on the basis of receiving the message of general 1. The only thing he can do is sending a messenger to general 1, acknowledging that he received the message and achieving that $K_1(K_2P)$. However, even if general 1 receives this acknowledgment, he is in a similar position as general 2 was in when he received the original message. Now general 2 does not know that the acknowledgment was delivered: $\neg K_2(K_1(K_2P))$. Because general 2 knows that without receiving the acknowledgment general 1 will not attack, general 2 cannot attack as long as he considers it possible that general 1 did not receive the acknowledgment. So, general 1 should send a message to general 2 in order to achieve that $K_2(K_1(K_2P))$. However, the problem now is that $\neg K_1(K_2(K_1(K_2P)))$, and so on. It turns out that no number of successful deliveries of acknowledgments can allow the generals to attack. Notice that, even if all the acknowledgments sent are received, *common knowledge* of P and hence coordination is not achieved, because of the uncertainty about what might have happened with the messengers.

Given a set $N = \{1, 2\}$ of agents (persons, computers) and a formula A , we may define the the notions of ‘everyone knows A ’ and ‘ A is common knowledge’.

Definition 6.16 (Common Knowledge). $EA := K_1A \wedge K_2A$ (everybody knows A); $E^0A := A$ and for $k = 0, 1, \dots$, $E^{k+1}A := E(E^kA)$. In particular, $E^1A = E(E^0A) = K_1A \wedge K_2A$ and $E^2A = E(E^1A) = K_1(K_1A \wedge K_2A) \wedge K_2(K_1A \wedge K_2A)$, which in $S4$ and $S5$ is equivalent to $K_1A \wedge K_1(K_2A) \wedge K_2(K_1A) \wedge K_2A$.

$CA := A \wedge EA \wedge E^2A \wedge E^3A \wedge \dots$ (A is *common knowledge*).

Notice that strictly speaking CA is not a formula in our language, because it is an infinite conjunction. For the Kripke semantics and the syntax (axiom and rule) of common knowledge see Fagin, e.a. [10] and Meyer and van der Hoek [29].

6.7 Completeness of Modal Propositional Logic

Let $K-$ be any of the modal systems K , KT or $KT4 = S4$. We shall prove completeness of modal logic, i.e., that any valid consequence in $K-$ of given premisses may be logically deduced by the tableaux rules of $K-$ from those premisses:

if $A_1, \dots, A_n \models B$ in $K-$, then $A_1, \dots, A_n \vdash' B$ in $K-$ (Theorem 6.7). (1)

We shall also prove:

if $A_1, \dots, A_n \vdash' B$ in $K-$, then $A_1, \dots, A_n \vdash B$ in $K-$ (Theorem 6.9). (2)

In Theorem 6.3 we have already shown the soundness of modal logic:

if $A_1, \dots, A_n \vdash B$ in $K-$, then $A_1, \dots, A_n \models B$ in $K-$. (3)

From (1), (2) and (3) it follows that the three notions $A_1, \dots, A_n \vdash B$ in $K-$, $A_1, \dots, A_n \models B$ in $K-$, and $A_1, \dots, A_n \vdash' B$ in $K-$ are equivalent.

In order to prove completeness of modal logic, we define a *procedure to construct a counterexample* to a given conjecture that $A_1, \dots, A_n \vdash' B$ in $K-$ with the following property: if the procedure fails, i.e., does not yield a counterexample, we have in fact constructed a tableau-deduction of B from A_1, \dots, A_n in $K-$. The procedure makes use of the tableaux rules and produces ‘trees’ which we shall call *search trees*.

Definition 6.17 (Procedure to construct a counterexample). In order to construct a counterexample to the conjecture that $A_1, \dots, A_n \vdash' B$ in $K-$, we must construct a Kripke model M for $K-$ such that for some world w in M , $M, w \models A_1 \wedge \dots \wedge A_n$, but $M, w \not\models B$.

Step 1: Start with $\{TA_1, \dots, TA_n, FB\}$ and apply all tableaux rules for the propositional connectives and the $T\Box$ rule in $K-$ as frequently as possible. However, in case one of the split-rules $T\rightarrow$, $T\vee$ and $F\wedge$ is applied, we make two search trees: one with the left split and one with the right split. Notice that for a tableau-deduction both search trees have to close.

For instance, consider the conjecture $\Diamond P \vdash' \Box \Diamond P \wedge \Diamond \Diamond P$ in KT :

search tree (1)	search tree (2)
$T \Diamond P, F \Box \Diamond P \wedge \Diamond \Diamond P$	$T \Diamond P, F \Box \Diamond P \wedge \Diamond \Diamond P$
$T \neg \Box \neg P, F \Box \Diamond P \wedge \Diamond \Diamond P$	$T \neg \Box \neg P, F \Box \Diamond P \wedge \Diamond \Diamond P$
$F \Box \neg P, F \Box \Diamond P \wedge \Diamond \Diamond P$	$F \Box \neg P, F \Box \Diamond P \wedge \Diamond \Diamond P$
$F \Box \neg P, F \Box \Diamond P$	$F \Box \neg P, F \Diamond \Diamond P$
	$F \Box \neg P, F \neg \Box \neg \neg \Box \neg P$
	$F \Box \neg P, T \Box \Box \neg P$
	$F \Box \neg P, T \Box \Box \neg P, T \Box \neg P$
	$F \Box \neg P, T \Box \Box \neg P, T \Box \neg P, T \neg P$
	$F \Box \neg P, T \Box \Box \neg P, T \Box \neg P, T \neg P, FP$

In the transition of the third line to the fourth line we apply the rule $F\wedge$ to $F \Box \Diamond P \wedge \Diamond \Diamond P$, which causes a split. At that stage we make two search trees, one with the left split signed formula $F \Box \Diamond P$ and one with the right split signed formula $F \Diamond \Diamond P$. One continues to apply all possible rules, except the $F\Box$ rule, as frequently as possible.

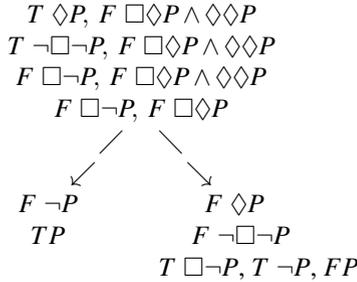
At this stage we have partially constructed one, two (or more) search trees, each consisting of one node labeled with signed formulas. A labeled node w in which all tableaux rules except the $F\Box$ -rule have been applied as frequently as possible will be called *logically complete*. Intuitively, this means that one has fully described which formulas are true and which formulas are false in the present world w . Next we continue to expand each search tree by one or more applications of the $F\Box$ rule.

Step 2 Each labeled node w in a search tree τ which is logically complete may contain one or more signed formulas of the form $F \Box A$. For each of the signed formulas of the form $F \Box A$ in a labeled node w we construct a new node w' , declare w' accessible from w in the given search tree τ , i.e., $wR_\tau w'$, and label this node w' with the formulas S_\Box , FA or $S_{T\Box}$, FA which result from applying the rule $F\Box$ to S , $F \Box A$ in K , KT or $S4$, respectively. Notice that formulas that occur in labeled

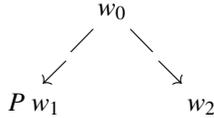
node w may not occur anymore in node w' and that for closure it suffices that at least one of the successor nodes contains TA and FA for some formula A .

Next we apply step 1 again, but now starting with S_{\square} , FA or $S_{T\square}$, FA , depending on the system K , KT or $S4$, resulting in one or more logically complete nodes (worlds) w' . Step 1 and 2 are repeated as frequently as possible.

For search-tree (1) above one can apply the $F\square$ rule to $F\square\neg P$, losing the $F\square\diamond P$ signed formula, and we can apply the $F\diamond$ rule to $F\square\diamond P$, losing the $F\square\neg P$ signed formula. For a tableau-deduction only one of these two options has to yield closure. So, we have two options to go on with search tree (1):

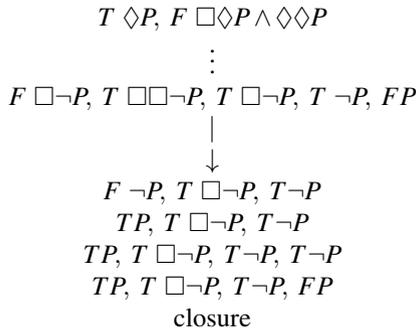


Whatever we do, we do not get closure. However, the nice thing is that we have constructed a search tree τ , starting with $T \diamond P, F \square \diamond P \wedge \diamond \diamond P$, in this case consisting of three nodes labeled with signed formulas, which yields a Kripke counterexample $M = \langle \{w_0, w_1, w_2\}, R_{\tau}, \models \rangle$ to the conjecture that $\diamond P \vdash' \square \diamond P \wedge \diamond \diamond P$ in KT :



By definition, $w_0 R_{\tau} w_1$, $w_0 R_{\tau} w_2$, $w_1 \models P$, corresponding with the occurrence of TP in node w_1 and $w_2 \not\models P$, corresponding with the occurrence of FP in node w_2 . One easily verifies that $M, w_0 \models \diamond P$, because $M, w_1 \models P$, but $M, w_0 \not\models \square \diamond P$ and hence $M, w_0 \not\models \square \diamond P \wedge \diamond \diamond P$, because $M, w_2 \not\models \diamond P$.

For search tree (2) there is only one formula of the form $F \square A$ in the upper node. Application of Step 2 results in the following search tree in KT , consisting of two nodes:



However, because search tree (1) does not close, we have not found a tableau-
deduction of $\Box\Diamond P \wedge \Diamond\Diamond P$ from $\Diamond P$ in KT . Instead, search tree (1) did not close
and yielded a Kripke counterexample to the conjecture $\Diamond P \vdash \Box\Diamond P \wedge \Diamond\Diamond P$ in KT .
In our example, after executing step 1, 2 and 1 once more, the two search trees are
finished and cannot be extended anymore.

Definition 6.18 (Search tree).

A search tree τ for the conjecture $A_1, \dots, A_n \vdash B$ in $K-$ is a set of nodes, labeled
with signed formulas, with a relation R_τ between the nodes, such that:

0. The upper node contains TA_1, \dots, TA_n, FB .

1. In case of K , $wR_\tau w' := w'$ is an immediate successor of w , i.e., w' results from the
application of the $F\Box$ rule to a formula of the form $F\Box A$ in w .

In case of KT , $wR_\tau w' := w = w'$ or w' is an immediate successor of w .

In case of $KT4 = S4$, $wR_\tau w' := w = w'$ or w' is a (not necessarily immediate) suc-
cessor of w .

2. For each node w in the search tree τ :

a) if $F C \rightarrow D$ occurs in w , then TC occurs in w **and** FD occurs in w ;

b) if $T C \wedge D$ occurs in w , then TC occurs in w **and** TD occurs in w ;

c) if $F C \vee D$ occurs in w , then FC occurs in w **and** FD occurs in w ;

d) if $T \neg C$ occurs in w , then FC occurs in w ;

e) if $F \neg C$ occurs in w , then TC occurs in w .

3. For each node w in the search tree τ :

a) if $T C \rightarrow D$ occurs in w , then FC occurs in w **or** TD occurs in w ;

b) if $F C \wedge D$ occurs in w , then FC occurs in w **or** FD occurs in w ;

c) if $T C \vee D$ occurs in w , then TC occurs in w **or** TD occurs in w .

4. For each node w in the search tree τ :

a) if $T \Box C$ occurs in w , then for all w' in τ with $wR_\tau w'$, TC occurs in w' ;

b) if $F \Box C$ occurs in w , then for some w' in τ with $wR_\tau w'$, FC occurs in w' .

Definition 6.19 (Closed/open search tree).

A search tree τ is *closed* if it contains at least one node labeled with TA and FA for
some formula A . Otherwise, the search tree is called *open*.

Theorem 6.5. *Let τ be an open search tree for the conjecture $A_1, \dots, A_n \vdash B$ in
 $K-$ with upper node w_0 . Let W_τ the set of nodes in τ and let R_τ be defined as
in Definition 6.18. Define $w \models P := TP$ occurs in w . Then $M_\tau = \langle W_\tau, R_\tau, \models \rangle$ is a
Kripke countermodel in $K-$ to the conjecture that $A_1, \dots, A_n \vdash B$. More precisely,
 $M_\tau, w_0 \models A_1 \wedge \dots \wedge A_n$, but $M_\tau, w_0 \not\models B$.*

Proof. Let τ be an open search tree with w_0 as upper node, containing TA_1, \dots, TA_n ,
 FB . Let $M_\tau = \langle W_\tau, R_\tau, \models \rangle$ be the corresponding Kripke model, as defined in the
theorem. We shall prove by induction:

1) If TA occurs in w , then $M_\tau, w \models A$.

2) If FA occurs in w , then $M_\tau, w \not\models A$.

Since TA_1, \dots, TA_n, FB occur in the top node w_0 , it follows that $M_\tau, w_0 \models A_1 \wedge \dots \wedge$
 A_n , but $M_\tau, w_0 \not\models B$. Therefore, $A_1, \dots, A_n \not\models B$ in $K-$.

Induction basis Let $A = P$ be atomic. If TP occurs in w , then by definition $w \models P$,

i.e., $M_\tau, w \models P$. If FP occurs in w , then - since τ is open - TP does not occur in w and hence by definition $w \not\models P$, i.e., $M_\tau, w \not\models P$.

Induction step Suppose 1) and 2) hold for C and D (induction hypothesis). We shall prove that 1) and 2) hold for $C \rightarrow D$, $C \wedge D$, $C \vee D$, $\neg C$ and $\Box C$.

Let $A = C \rightarrow D$ and suppose $TC \rightarrow D$ occurs in w . Then according to Definition 6.18, 3 a), FC is in w or TD is in w . So, by the induction hypothesis, $M_\tau, w \not\models C$ or $M_\tau, w \models D$. Consequently, $M_\tau, w \models C \rightarrow D$.

Let $A = C \rightarrow D$ and suppose $FC \rightarrow D$ occurs in w . Then according to Definition 6.18, 2 a), TC is in w and FD is in w . So, by the induction hypothesis, $M_\tau, w \models C$ and $M_\tau, w \not\models D$. Consequently, $M_\tau, w \not\models C \rightarrow D$.

The cases that $A = C \wedge D$, $A = C \vee D$ and $A = \neg C$ are treated similarly.

Let $A = \Box C$ and suppose $T\Box C$ occurs in w . Then according to Definition 6.18, 4 a), for every node w' in τ with $wR_\tau w'$, TC occurs in w' . So, by the induction hypothesis, for all w' in τ , if $wR_\tau w'$, then $M_\tau, w' \models C$ and hence $M_\tau, w \models \Box C$.

Let $A = \Box C$ and suppose $F\Box C$ occurs in w . Then according to Definition 6.18, 4 b), there is a node w' in τ with $wR_\tau w'$ such that FC occurs in w' . So, by the induction hypothesis, $M_\tau, w' \not\models C$ and hence $M_\tau, w \not\models \Box C$. \square

Theorem 6.6. *If all search trees for the conjecture $A_1, \dots, A_n \vdash' B$ in $K-$ are closed, i.e., contain closure in one of their branches, then $A_1, \dots, A_n \vdash' B$ in $K-$.*

Proof. Suppose all search trees for the conjecture $A_1, \dots, A_n \vdash' B$ in $K-$ are closed. Then it follows from the construction of the search trees that the closed branches together form a tableau-deduction of B from A_1, \dots, A_n in $K-$. \square

Example 6.10. We construct the search trees for the conjecture $\Diamond(P \wedge Q) \vdash' \Diamond P \wedge (\Diamond Q \vee \Box P)$ in K . Step 1 yields two partial search trees each consisting of one node:

$$\begin{array}{ll} T \Diamond(P \wedge Q), F \Diamond P \wedge (\Diamond Q \vee \Box P) & T \Diamond(P \wedge Q), F \Diamond P \wedge (\Diamond Q \vee \Box P) \\ T \Diamond(P \wedge Q), F \Diamond P & T \Diamond(P \wedge Q), F \Diamond Q \vee \Box P \\ T \neg \Box \neg (P \wedge Q), F \neg \Box \neg P & T \neg \Box \neg (P \wedge Q), F \neg \Box \neg Q, F \Box P \\ F \Box \neg (P \wedge Q), T \Box \neg P & F \Box \neg (P \wedge Q), T \Box \neg Q, F \Box P \end{array}$$

Because there is no $T\Box$ rule for K , step 1 finishes here. The only rule which may be applied next is the rule $F\Box$ for K . Applying step 2 to the last sequents of step 1 we get:

$$\begin{array}{ll} F \Box \neg (P \wedge Q), T \Box \neg P & F \Box \neg (P \wedge Q), T \Box \neg Q, F \Box P \\ \downarrow & \swarrow \quad \searrow \\ F \neg (P \wedge Q), T \neg P & F \neg (P \wedge Q), T \neg Q & T \neg Q, FP \\ T P \wedge Q, FP & TP \wedge Q, FQ & FQ, FP \\ TP, TQ, FP & TP, TQ, FQ & \end{array}$$

The leftmost search tree consists of one branch with two nodes, and is closed. The rightmost search tree consists of two branches and three nodes; its left branch is closed and its right branch is open. The two closed branches together form a tableau-deduction in K of $\Diamond P \wedge (\Diamond Q \vee \Box P)$ from $\Diamond(P \wedge Q)$.

Theorem 6.7 (Completeness).

If $A_1, \dots, A_n \models B$ in $K-$, then $A_1, \dots, A_n \vdash' B$ in $K-$.

Proof. Suppose $A_1, \dots, A_n \models B$ in $K-$. Construct all search trees for the conjecture $A_1, \dots, A_n \vdash' B$ in $K-$. If one of them is open, say τ , then by Theorem 6.5, $M_{\tau, w_0} \models A_1 \wedge \dots \wedge A_n$, while $M_{\tau, w_0} \not\models B$. This contradicts the assumption $A_1, \dots, A_n \models B$ in $K-$. Hence, there can be no open search tree for the conjecture $A_1, \dots, A_n \vdash' B$ in $K-$. That is, all search trees for this conjecture are closed. So, by Theorem 6.6, $A_1, \dots, A_n \vdash' B$ in $K-$. \square

In the case of K , resp. KT , our procedure to construct a counterexample to the conjecture $A_1, \dots, A_n \vdash' B$ will stop after finitely many steps and then either yield a Kripke counterexample or a tableau-deduction of B from A_1, \dots, A_n in K , resp. KT . In the case of $S4$, this procedure does not necessarily stop after finitely many steps (see Example 6.9), but nevertheless after finitely many steps it will become clear whether one has constructed a Kripke counterexample in $S4$ or a tableau-deduction of B from A_1, \dots, A_n in $S4$. Therefore, the modal propositional logics K , KT and $S4$ are *decidable*.

Theorem 6.8 (Decidability). *The modal propositional logics K , KT and $S4$ are decidable, i.e., there is a procedure to decide whether $A_1, \dots, A_n \vdash' B$ in K , KT , resp. $S4$, in finitely many steps.*

In order to prove that the three notions of formal deducibility in $K-$, Kripke valid consequence in $K-$ and tableau-deducibility in $K-$ are equivalent we still have to show the following theorem.

Theorem 6.9. *If $A_1, \dots, A_n \vdash' B$ in $K-$, then $A_1, \dots, A_n \vdash B$ in $K-$.*

Proof. The proof is a generalization of the analogue for classical propositional logic; see Theorem 2.27. Suppose $A_1, \dots, A_n \vdash' B$ in $K-$, i.e., B is tableau-deducible from A_1, \dots, A_n in $K-$. It suffices to show:

for every sequent $S = \{TD_1, \dots, TD_k, FE_1, \dots, FE_m\}$ in a tableau-deduction of B from A_1, \dots, A_n in $K-$ it holds that $D_1, \dots, D_k \vdash E_1 \vee \dots \vee E_m$ in $K-$. (*)

Consequently, because $\{TA_1, \dots, TA_n, FB\}$ is the first (upper) sequent in any given tableau-deduction of B from A_1, \dots, A_n in $K-$, it follows that $A_1, \dots, A_n \vdash B$ in $K-$.

The proof of (*) is tedious, but has a simple plan: the statement is true for the final sequents in a tableau-deduction in $K-$, and the statement remains true if we go up in the tableau-deduction in $K-$ via the T and F rules.

Basic step: Any final sequent in a tableau-deduction of B from A_1, \dots, A_n in $K-$ is of the form $\{TD_1, \dots, TD_k, TP, FP, FE_1, \dots, FE_m\}$. So, we have to show that $D_1, \dots, D_k, P \vdash P \vee E_1 \vee \dots \vee E_m$. And this is straightforward: $D_1, \dots, D_k, P \vdash P$ and $P \vdash P \vee E_1 \vee \dots \vee E_m$.

Induction step: We have to show that for all rules of $K-$ the following is the case: if (*) holds for all lower sequents in the rule (induction hypothesis), then (*) holds for the upper sequent in the rule.

In the proof of Theorem 2.27 we have already shown the induction step for the T - and F -rules for the connectives. So, we may restrict ourselves to the T - and F -rules for \square in system $K-$.

Induction step for rule $F\Box$ in K : For convenience, we will suppose that $S = \{T\Box C, TD, FE\}$. So, consider:
$$\frac{T\Box C, TD, FE, F\Box A}{TC, FA}$$

By the induction hypothesis, we have $C \vdash A$ in K . We have to show: $\Box C, D \vdash E \vee \Box A$ in K . This is straightforward: from $C \vdash A$ in K follows $\Box C \vdash \Box A$ in K and hence, $\Box C, D \vdash E \vee \Box A$ in K .

Induction step for rule $T\Box$ in KT : For convenience, we will suppose that $S = \{TD, FE\}$. So, consider:
$$\frac{T\Box C, TD, FE}{TC, TD, FE}$$

By the induction hypothesis, we have $C, D \vdash E$ in KT . We have to show: $\Box C, D \vdash E$ in KT . This is straightforward, because $\Box C \rightarrow C$ is an axiom of KT .

The other T - and F -rules for \Box in K – are treated similarly. □

Exercise 6.17. Construct a counterexample showing that the cosmological proof of God’s existence in $S5$, given in Exercise 6.3, does not hold in $S4$:

$\Diamond P, \Box(\Diamond P \rightarrow Q) \not\vdash \Box Q$ in $S4$.

Exercise 6.18. Construct a counterexample showing that the ontological proof of God’s existence in $S5$, given in Exercise 6.4, does not hold in $S4$:

$\Box(Q \rightarrow \Box Q), \Diamond Q \not\vdash Q$ in $S4$.

Exercise 6.19. Prove or refute: a) $\Box(S \rightarrow E), \Box(E \rightarrow L), \neg\Diamond L \vdash \neg\Diamond S$ in K .

b) $S \rightarrow \Box E, E \rightarrow \Box L, \neg\Diamond L \vdash \neg\Diamond S$ in $S4$ (confer Exercise 6.1).

6.8 Strict Implication

The material implication, \rightarrow , of classical propositional logic is characterized in terms of its truth table: $P \rightarrow Q$ is 0 (false) if and only if P is 1 (true) and Q is 0 (false). Through the ages objections have been raised against the ‘only if’: if P is 0, then $P \rightarrow Q$ is 1. Although there are many arguments in favor of the truth table of $P \rightarrow Q$, as we have seen in Section 2.2, also objections have been raised, in particular the so-called *paradoxes of material implication*:

a) $\neg A \models A \rightarrow B$: if A is false, then from A follows any proposition B ;

b) $B \models A \rightarrow B$: if B is true, then B follows from any proposition A .

So, from ‘I do not break my leg’ it logically follows that ‘if I break my leg, then I go for skyng’ and from ‘I like my coffee’ it logically follows that ‘if there is oil in my coffee, then I like my coffee’; see Section 2.4. In the same section we have seen that P. Grice [16] explains these paradoxes by pointing out that one should take into account not only the truth conditions of the propositions asserted, but also the pragmatic principles governing discourse: $A \rightarrow B$ is normally not to be asserted by someone who is in the position to deny A or to assert B .

The dispute between advocates of the truth-functional account of conditionals, given in Section 2.2, and the advocates of other - more complex but seemingly more adequate - accounts is as old as logic itself. The truth-functional account is first

known to have been proposed by Philo of Megara ca. 300 B.C. in opposition to the view of his teacher Diodorus Cronus. We know of this through the writings of Sextus Empiricus some 500 years later, the earlier documents having been lost; see Section 2.10.2. Sextus reports Philo as attributing truth values to conditionals just as in our truth table for \rightarrow . Diodorus probably had in mind what later was called *strict implication*.

Rejecting material implication as an adequate representation of ‘if ... , then ...’, in 1918 C.I. Lewis [25] put forward *strict implication*, \mapsto , which can be rendered in terms of necessity and material implication: $\Box(A \rightarrow B)$.

Definition 6.20. *Strict implication*, \mapsto , is defined by $A \mapsto B := \Box(A \rightarrow B)$.

It is easy to show that the versions for strict implication of the paradoxes of material implication do not hold. According to Exercise 6.20:

a) not $\neg A \vdash' A \mapsto B$ in $S4$; and b) not $B \vdash' A \mapsto B$ in $S4$.

However, the definition of strict implication leads to the so-called *paradoxes of strict implication*. According to Exercise 6.21:

- a) $\Box \neg A \vdash' A \mapsto B$ in K : an impossible proposition A implies every proposition B .
- b) $\Box B \vdash' A \mapsto B$ in K : a necessary proposition B is implied by every proposition A .
- c) $Q \vdash' P \mapsto P$ in K and d) $\vdash' \neg Q \wedge Q \mapsto P$ in K .

The problem with these paradoxes is that for the provability of an inference from A to B , A should be *relevant* to B . See Section 6.10.

Exercise 6.20. Prove: not $\neg A \vdash' A \mapsto B$ in $S4$ and not $B \vdash' A \mapsto B$ in $S4$.

Exercise 6.21. Prove the following so-called *paradoxes of strict implication*:

- a) $\Box \neg A \vdash' A \mapsto B$ in K ; b) $\Box B \vdash' A \mapsto B$ in K ;
- c) $Q \vdash' P \mapsto P$ in K ; d) $\vdash' \neg Q \wedge Q \mapsto P$ in K .

6.9 Counterfactuals

Counterfactuals are expressions of the form $A \Box \rightarrow B$, to be read as ‘if it were the case that A , then it would be the case that B ’, where A is supposed to be false. Unlike material, strict and relevant implication, the counterfactual

a) is not transitive, i.e., not $\frac{A \Box \rightarrow B \quad B \Box \rightarrow C}{A \Box \rightarrow C}$,

b) does not have the property of contraposition $\frac{A \Box \rightarrow B}{\neg B \Box \rightarrow \neg A}$, and

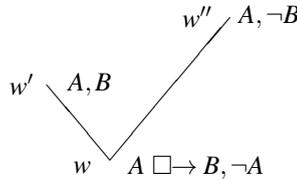
c) does not have the property of strengthening $\frac{A \Box \rightarrow B}{A \wedge C \Box \rightarrow B}$.

The following counterexamples are from D. Lewis [26]:

a) If J. Edgar Hoover had been born a Russian, then he would have been a communist. If he had been a communist, he would have been a traitor.

Therefore: If he had been born a Russian, he would have been a traitor.

- b) If Boris had gone to the party, Olga would still have gone.
 Therefore: If Olga had not gone, Boris would still not have gone.
 Suppose that Boris wanted to go, but stayed away solely in order to avoid Olga, so the conclusion is false; but Olga would have gone all the more willingly if Boris had been there, so the premiss is true.
- c) If I walked on the lawn, no harm at all would come of it. Therefore:
 If I and everyone else walked on the lawn, no harm at all would come of it.



We say that $A \Box \rightarrow B$ is true in world w iff either A is impossible in w or there is an accessible $A \wedge B$ -world w' , which is *closer to* w than every $A \wedge \neg B$ -world is (R. Stalnaker, D. Lewis, \pm 1970), where a C -world is simply a world in which C is true.

- Example 6.11.* a) A young child to his father: If you would bring that big tree home (A), I would make matches from it (B). This proposition is true in the present world because the child considers the antecedent A to be impossible.
- b) If you would jump out of the window at the 20th floor (A), you would get injured (B). This proposition is true in the present world w , because there is world w' in which $A \wedge B$ is true and which is closer to w than each world w'' in which $A \wedge \neg B$ is true.
- c) If you would jump out of the window at the 20th floor, you would change into a bird. This proposition is not true in the present world w , because we cannot imagine a world w' in which $A \wedge B$ is true and which is closer to w than any world in which $A \wedge \neg B$ is true.

Given a Kripke model $M = \langle W, R, \models \rangle$, we assume that for each w in W there is a binary relation $<_w$ on W , where $w' <_w w''$ is to be read as: w' is *closer to* w than w'' . Furthermore, we assume that R is reflexive, and

1. if wRw' and not wRw'' , then $w' <_w w''$;
2. for all w, w' in W , if $w \neq w'$, then $w <_w w'$ and not $w' <_w w$.

Definition 6.21 ($M \models_w A \Box \rightarrow B$). Let $M = \langle W, R, \models, < \rangle$ be a Kripke model, where for each w in W , $<_w$ is a binary relation on W , satisfying the conditions just mentioned. $M \models_w A \Box \rightarrow B := M \models_w \neg \Diamond A$ or there is some world w' in W such that a) wRw' and $M \models_{w'} A \wedge B$, and b) for all w'' in W , if $M \models_{w''} A \wedge \neg B$, then $w' <_w w''$.

For an illustration we refer to Exercise 6.22.

Under the conditions just mentioned, counterfactuals with true antecedents reduce to material conditionals. More precisely, the following two inference-patterns are valid:

$$(a) \frac{A \wedge \neg B}{\neg(A \Box \rightarrow B)} \text{ and } (b) \frac{A \wedge B}{A \Box \rightarrow B};$$

that is, our truth conditions guarantee that whenever the premiss is true in a world of a given model M , then so is the conclusion; see Exercise 6.23.

The validity of the first inference-pattern (a) also guarantees the validity of the inference from a counterfactual to a material conditional and the validity of Modus Ponens for a counterfactual conditional: $\frac{A \Box \rightarrow B}{A \rightarrow B}$ and $\frac{A \quad A \Box \rightarrow B}{B}$. We also have the inference: $\frac{\Box(A \rightarrow B)}{A \Box \rightarrow B}$; see Exercise 6.24.

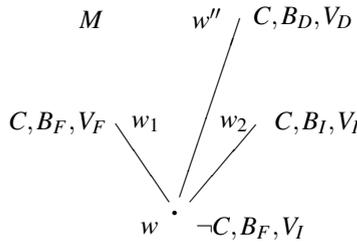
One can develop possible-world semantics for counterfactuals, a notion of validity ($\models A$) and a notion of provability ($\vdash A$) such that a counterfactual formula A is valid if and only if A is provable. See D. Lewis' paper [27], pp. 441-443, or his monograph [26]; de Swart [37] and Gent [14].

Let $A \Diamond \rightarrow B$ stand for 'if A were the case, then B might be the case'. Then it is plausible to have $A \Diamond \rightarrow B$ iff $\neg(A \Box \rightarrow B)$. The reader can check for himself that, given plausible assumptions about comparative similarity of worlds where Bizet and Verdi would be compatriots, both

- (1) if Bizet and Verdi were compatriots, then Bizet might be Italian, and
 - (2) If Bizet and Verdi were compatriots, then Bizet might not be Italian,
- are true. For further reading see Harper e.a. [17].

Exercise 6.22. Let C stand for: Bizet and Verdi are compatriots. Let B_F, B_I and B_D stand for: Bizet is French, Italian, Dutch, respectively. And similarly, V_F, V_I and V_D for: Verdi is French, Italian, Dutch, respectively. Let w be the actual world, in which B_F and V_I hold, of the following Kripke model M . Verify that in the Stalnaker-Lewis analysis of counterfactuals:

- a) $M, w \models C \Box \rightarrow (B_F \wedge V_F) \vee (B_I \wedge V_I)$
- b) $M, w \not\models C \Box \rightarrow B_I$,
- c) $M, w \not\models C \Box \rightarrow V_F$
- d) $M, w \models C \Box \rightarrow \neg B_D \wedge \neg V_D$.



Exercise 6.23. Let $M = \langle W, R, \models \rangle$ be a Kripke model with R reflexive and for each w in W , let $<_w$ be a binary relation on W satisfying: 1. if wRw' and not wRw'' , then $w' <_w w''$ and 2. if $w \neq w'$, then $w <_w w'$ and not $w' <_w w$. Prove:

- a) if $M, w \models A \wedge \neg B$, then $M, w \models \neg(A \Box \rightarrow B)$, and
- b) if $M, w \models A \wedge B$, then $M, w \models A \Box \rightarrow B$.

Exercise 6.24. Under the conditions mentioned in Exercise 6.23 prove that: if $M, w \models \Box(A \rightarrow B)$, then $M, w \models A \Box \rightarrow B$.

Exercise 6.25. Show that $\text{not} \models (P \mapsto Q) \vee (Q \mapsto P)$ in $S4$ and that also $\text{not} \models (P \Box \rightarrow Q) \vee (Q \Box \rightarrow P)$, while $\models (P \rightarrow Q) \vee (Q \rightarrow P)$.

6.10 Weak and Relevant Implication; Entailment*

In his paper *The weak theory of implication*, A. Church [8] succeeds in excluding the paradoxes of strict implication (see Exercise 6.21) without also excluding at the same time arguments which everyone regards as valid. In his paper A. Church presents essentially the following axiom schemes for what he calls *weak implication*, but what one might also call *relevant implication*, and which we denote by \Rightarrow :

1. $A \Rightarrow A$
2. $(A \Rightarrow B) \Rightarrow ((B \Rightarrow C) \Rightarrow (A \Rightarrow C))$
3. $(A \Rightarrow (B \Rightarrow C)) \Rightarrow ((A \Rightarrow B) \Rightarrow (A \Rightarrow C))$
4. $(A \Rightarrow (B \Rightarrow C)) \Rightarrow (B \Rightarrow (A \Rightarrow C))$

together with the rule Modus Ponens, $\frac{B \quad B \Rightarrow C}{C}$.

\Rightarrow satisfies principles of relevance in the following mathematically definite sense:

$$A_1, \dots, A_{n-1} \vdash^* A_n \Rightarrow B \text{ iff } A_1, \dots, A_{n-1}, A_n \vdash^* B,$$

where $A_1, \dots, A_{n-1}, A_n \vdash^* B$ (B is deducible from A_1, \dots, A_n) means that B can be obtained by a finite number of applications of Modus Ponens to A_1, \dots, A_{n-1}, A_n and to instances of the axiom schemes 1, 2, 3, 4, *such that all of A_1, \dots, A_{n-1}, A_n actually are used in the deduction of B* ; more precisely, such that B gets the relevance-index $\{1, \dots, n-1, n\}$ if we assign to each A_i ($1 \leq i \leq n$) the index $\{i\}$ and to each consequence of an application of Modus Ponens the union of the indices of its premisses.

For instance, $A \Rightarrow B, B \Rightarrow C \vdash^* A \Rightarrow C$, for the following schema is a deduction (in the new sense) of $A \Rightarrow C$ from $A \Rightarrow B$ and $B \Rightarrow C$:

$$\frac{\begin{array}{c} A \Rightarrow B_{\{1\}} \qquad (A \Rightarrow B) \Rightarrow ((B \Rightarrow C)) \Rightarrow (A \Rightarrow C) \\ \hline B \Rightarrow C_{\{2\}} \qquad ((B \Rightarrow C) \Rightarrow (A \Rightarrow C))_{\{1\}} \\ \hline (A \Rightarrow C)_{\{1,2\}} \end{array}}$$

However, it is not the case that $Q \vdash^* P \Rightarrow P$ (see Anderson & Belnap, [2]), while $Q \vdash P \rightarrow P$ does hold, since in ' $A \vdash B$ ' it is not demanded that A actually is used in the deduction of B and $\vdash P \rightarrow P$ holds; see Section 2.6.

We define $M = \langle S, \emptyset, \cup, \models \rangle$ to be a *model* (for the logic of weak or relevant implication) if and only if

1. S is a collection of sets, closed under \cup ; the elements of S are to be regarded as pieces of information;
2. \emptyset is the empty set (regarded as the empty piece of information);

3. $a \cup b$ is the union of a and b (see Chapter 3);
4. \models is a relation between elements of S and atomic formulas P ; ' $a \models P$ ' is to be read as: P is true on the basis of the information in a .

For M a model and A a formula built from atomic formulas by means of \Rightarrow only, we define $M \models_a A$ (A is true on the basis of the information a of the model M) as follows:

$M \models_a P$ iff $a \models P$ (P atomic);

$M \models_a B \Rightarrow C$ iff for all b in S , not $M \models_b B$ or $M \models_{a \cup b} C$.

M is a model for A ($M \models^* A$) iff $M \models_{\emptyset} A$.

A is *valid* ($\models^* A$) iff for all models M , $M \models^* A$. And B is a *valid consequence* of A_1, \dots, A_n ($A_1, \dots, A_n \models^* B$) iff $\models^* A_1 \Rightarrow (A_2 \Rightarrow \dots (A_n \Rightarrow B) \dots)$.

Exercise 6.26. Prove the (deduction) theorem: $A_1, \dots, A_{n-1} \vdash^* A_n \Rightarrow B$ iff $A_1, \dots, A_{n-1}, A_n \vdash^* B$. Hint: the proof of the 'if'-part of the (deduction) theorem proceeds by replacing in a given deduction of B from A_1, \dots, A_{n-1}, A_n each expression C_c with $n \in c$ by $(A_n \Rightarrow C)_{c-\{n\}}$. For Modus Ponens, $\frac{D_d \quad (D \Rightarrow E)_e}{E_{d \cup e}}$, four different cases arise, depending on whether $n \in d$ and/or $n \in e$; and the axioms have been chosen such that the resulting schema can easily be supplemented to a deduction of $A_n \Rightarrow B$ from A_1, \dots, A_{n-1} .

Exercise 6.27. a) Prove the *Soundness Theorem*:

if $A_1, \dots, A_n \vdash^* B$, then $A_1, \dots, A_n \models^* B$.

In [38] A. Urquhart also proves the converse of this statement, i.e., completeness.

b) Prove that $\not\models^* Q \Rightarrow (P \Rightarrow P)$, and hence $Q \not\vdash^* P \Rightarrow P$. In general, the relevant implication versions of the original paradoxes of strict implication do not hold.

Exercise 6.28. Prove that $\vdash^* A \Rightarrow ((A \Rightarrow A) \Rightarrow A)$. This says that if A is true, then it follows from $A \Rightarrow A$. But it seems reasonable to suppose that any logical consequence of $A \Rightarrow A$ should necessarily be true (see Anderson & Belnap [2], p. 23). We therefore consider *entailment*, \twoheadrightarrow , defined by $P \twoheadrightarrow Q := \Box(P \Rightarrow Q)$, which was essentially considered for the first time by W. Ackermann [1] in his *Begründung einer strengen Implikation*. In this paper W. Ackermann presents essentially the following axiomatic system for \twoheadrightarrow :

1. $A \twoheadrightarrow A$
2. $(A \twoheadrightarrow B) \twoheadrightarrow ((B \twoheadrightarrow C) \twoheadrightarrow (A \twoheadrightarrow C))$
3. $(A \twoheadrightarrow (B \twoheadrightarrow C)) \twoheadrightarrow ((A \twoheadrightarrow B) \twoheadrightarrow (A \twoheadrightarrow C))$
4. $(A \twoheadrightarrow B) \twoheadrightarrow (((A \twoheadrightarrow B) \twoheadrightarrow C) \twoheadrightarrow C)$

together with the rule Modus Ponens $\frac{P \quad P \twoheadrightarrow Q}{Q}$.

Entailment satisfies both principles of relevance and principles of necessity in certain mathematically definite senses: all valid entailments are necessarily valid and in all valid entailments the antecedent is relevant to the succedent.

6.11 Modal Predicate Logic

A *possible world semantics* for the modal predicate logics is obtained by demanding that a Kripke model M contains for every world w in M a domain or universe $U(w)$ such that if wRw' , then $U(w)$ is a subset of $U(w')$.

$M, w \models \forall x[A(x)] :=$ for every individual d in $U(w)$, $M, w \models A(a)[d]$, and

$M, w \models \exists x[A(x)] :=$ there is some d in $U(w)$ such that $M, w \models A(a)[d]$.

'It is possible that unicorns exist' can be rendered by $\Diamond\exists x[P(x)]$, and is likely to be true. But 'there is an object which possibly is a unicorn', to be rendered by $\exists x[\Diamond P(x)]$, is generally held to be false.

In terms of possible worlds, the difference can be explained as follows, using $U(w)$ for the universe of world w :

$M, w \models \Diamond\exists x[P(x)] :=$ there is a world w' in M accessible from w (wRw') such that $M, w' \models \exists x[P(x)]$, i.e., there is a world w' in M accessible from w such that there is an individual d in the universe $U(w')$ of w' which is a unicorn in w' .

$M, w \models \exists x[\Diamond P(x)] :=$ there is an object d in the universe $U(w)$ of w such that $M, w \models \Diamond P(a)[d]$, i.e., there is an object d in the universe $U(w)$ of w such that there is a world w' in M accessible from w (wRw') in which d is a unicorn.

Supposing that if wRw' , then the universe $U(w)$ of w is a subset of the universe $U(w')$ of w' , we find that

$$\text{if } M, w \models \exists x[\Diamond P(x)], \text{ then } M, w \models \Diamond\exists x[P(x)],$$

but not conversely. Hence the following statements hold, but not conversely:

$$\begin{aligned} &\models \exists x[\Diamond\neg A(x)] \rightarrow \Diamond\exists x[\neg A(x)]. \\ &\models \neg\Diamond\exists x[\neg A(x)] \rightarrow \neg\exists x[\Diamond\neg A(x)]. \\ &\models \neg\neg\Box\neg\exists x[\neg A(x)] \rightarrow \neg\exists x[\neg\Box\neg\neg A(x)]. \\ &\models \Box\forall x[A(x)] \rightarrow \forall x[\Box A(x)]. \end{aligned}$$

Again, the difference between $\Box\forall x[A(x)]$ and $\forall x[\Box A(x)]$ may be explained best in terms of possible world semantics:

$M, w \models \Box\forall x[A(x)] :=$ for every world w' in M with wRw' and for every object d in the universe $U(w')$ of w' , $M, w' \models A(a)[d]$; but

$M, w \models \forall x[\Box A(x)] :=$ for every object d in the universe $U(w)$ of w and for every world w' in M with wRw' , $M, w' \models A(a)[d]$.

A *Hilbert-type proof system* for the modal predicate logics K , KT , $S4$, and $S5$, is obtained by adding to the axioms and rules for the respective modal propositional logics the (classical) axioms and rules for the quantifiers:

\forall axiom: $\forall x[A(x)] \rightarrow A(t)$ and \exists axiom: $A(t) \rightarrow \exists x[A(x)]$ for any term t .

\forall rule: from $C \rightarrow A(a)$ deduce $C \rightarrow \forall x[A(x)]$, provided a does not occur in C .

\exists rule: from $A(a) \rightarrow C$ deduce $\exists x[A(x)] \rightarrow C$, provided a does not occur in C .

Let us show that $\vdash \Box\forall x[A(x)] \rightarrow \forall x[\Box A(x)]$ in K :

1. $\forall x[A(x)] \rightarrow A(a)$ by the axiom for \forall .
2. $\Box(\forall x[A(x)] \rightarrow A(a))$ from 1 by the rule for \Box .

3. $\Box\forall x[A(x)] \rightarrow \Box A(a)$ from 2 and the axiom for \Box , using Modus Ponens.
4. $\Box\forall x[A(x)] \rightarrow \forall x[\Box A(x)]$ from 3 by the rule for \forall .

A *tableaux proof system* for the modal predicate logics K , KT and $S4$ is obtained by adding the T - and F -rules for \forall and \exists to the tableaux rules for the connectives and \Box :

$$\begin{array}{cc}
 S, T \forall x[A(x)] & S, F \forall x[A(x)] \\
 S, T \forall x[A(x)], TA(t) & S, FA(a) \text{ with } a \text{ new} \\
 \\
 S, T \exists x[A(x)] & S, F \exists x[A(x)] \\
 S, T A(a) \text{ with } a \text{ new} & S, F \exists x[A(x)], FA(t)
 \end{array}$$

Soundness and completeness of the modal predicate logics with respect to the appropriate Kripke semantics can again be shown by generalizing the proofs for the propositional case in Section 6.7.

Although $\Box\forall x[A(x)] \rightarrow \forall x[\Box A(x)]$ is formally provable in K and hence Kripke-valid, the converse formula $\forall x[\Box A(x)] \rightarrow \Box\forall x[A(x)]$, called the *Barcan formula*, is not Kripke-valid. A Kripke counterexample in $S4$ can be obtained by trying to construct a tableau-proof of this formula; we do not succeed in finding such a proof, but instead we find an open search tree from which we can immediately read off a counterexample.

$$\begin{array}{c}
 \begin{array}{c}
 T \forall x[\Box A(x)], F \forall x[A(x)] \\
 T \Box A(a_1), T \forall x[\Box A(x)], F \forall x[A(x)] \\
 TA(a_1), T \Box A(a_1), T \forall x[\Box A(x)], F \forall x[A(x)] \\
 \hline
 T \Box A(a_1), \quad F \forall x[A(x)] \\
 T \Box A(a_1), \quad FA(a_2) \\
 T \Box A(a_1), TA(a_1), \quad FA(a_2)
 \end{array}
 \quad
 \begin{array}{c}
 M \\
 \{a_1\} \quad | \quad A(a_1) \\
 \\
 \{a_1, a_2\} \quad | \quad A(a_1)
 \end{array}
 \end{array}$$

Let $M = \langle \{w_1, w_2\}, R, \models \rangle$ be the Kripke model in $S4$ consisting of two worlds w_1, w_2 with $w_1 R w_2$, $U(w_1) = \{a_1\}$, $U(w_2) = \{a_1, a_2\}$, $w_1 \models A(a_1)$ and $w_2 \models A(a_1)$, but $w_2 \not\models A(a_2)$, corresponding with the occurrence of $TA(a_1)$ in w_1 and in w_2 and the occurrence of $FA(a_2)$ in w_2 .

$M, w_1 \models \forall x[\Box A(x)] :=$ all objects in $U(w_1)$ have the property $\Box A$ in w_1 . This is the case, because $M, w_1 \models A(a_1)$ and $M, w_2 \models A(a_1)$. But $M, w_1 \models \Box\forall x[A(x)] :=$ for all worlds w' in M accessible from w_1 each individual in $U(w')$ has the property A in w' . Since $w_1 R w_2$, a_2 in $U(w_2)$ and by definition $w_2 \not\models A(a_2)$, it follows that $M, w_1 \not\models \Box\forall x[A(x)]$.

Exercise 6.29. Show that $\vdash' \Box\forall x[A(x)] \rightarrow \forall x[\Box A(x)]$ in K .

6.11.1 Modal Predicate Logic and Essentialism

Leibniz' law says that those things are the same of which one may be substituted for the other with preservation of truth. In contemporary treatments of identity this law is presented as follows:

$$(1) \models a = b \rightarrow (\dots a \dots \Leftrightarrow \dots b \dots)$$

where $\dots a \dots$ is a context containing occurrences of the name a , and $\dots b \dots$ is the same context except that one or more occurrences of a have been replaced by b : if $a = b$, then what holds for a also holds for b and vice versa.

In the propositional calculus we have a similar principle, called the replacement theorem:

$$(2) \models (A \Leftrightarrow B) \rightarrow (\dots A \dots \Leftrightarrow \dots B \dots).$$

And the analogue of the replacement theorem for predicate logic is principle

$$(3) \models (P(a) \Leftrightarrow Q(a)) \rightarrow (\dots P(a) \dots \Leftrightarrow \dots Q(a) \dots).$$

Quine [33] and Føllesdal [12] have argued that in order to make sense of quantified modal logic, modal contexts should be *referentially transparent*, i.e., principle (1) should hold also for modal contexts, and at the same time they should be *extensionally opaque*, i.e., the principles (2) and (3) should NOT hold for modal contexts.

According to Quine, in order to be able to quantify into modal contexts, these contexts should be referentially transparent: $\exists x[\Box(x > 7)]$ holds because $\Box(9 > 7)$ is true; but $9 =$ the number of planets (in this world); so, \Box (the number of planets (in this world) > 7) should hold.

Therefore, quantified modal logic only makes sense if we accept principle (1) also for modal contexts:

$$(1) \models a = b \rightarrow (\dots a \dots \Leftrightarrow \dots b \dots).$$

Principle (1) says that whatever is asserted to be true of an object, must be true of it regardless of how it is referred to. In other words, modal contexts should be *referentially transparent*, i.e., if two singular terms refer to the same object, they are interchangeable with preservation of truth (also in modal contexts).

Principle (1) says in particular that $\models a = b \rightarrow (\Box(a = a) \Leftrightarrow \Box(a = b))$. And since $\Box(a = a)$ is valid, it follows that

$$(1^*) : \models a = b \rightarrow \Box(a = b).$$

For instance, Hesperus and Phosphorus are two different names referring to the same object (the planet Venus), i.e., Hesperus = Phosphorus, and hence, $\Box(\text{Hesperus} = \text{Phosphorus})$.

Principle (1^{*}) says that if a and b refer to the same object, say o , in this world, then they refer to the same object (but possibly different from o) in any world accessible from this one. Hence, if a is a rigid designator (i.e., refers to the same object in any world accessible from this one), then b is also a rigid designator. In fact, Kripke already argued that proper names and nouns for natural kinds are rigid designators; see Subsection 6.6.2.

On the other hand, if we accept one of the principles (2) or (3) also for modal contexts (i.e., contexts $\dots A \dots$ or $\dots P(a) \dots$ containing modalities), then it even follows that $\models B \Leftrightarrow \Box B$ for any proposition B . In other words, the extension of (2) or (3) from classical propositional logic or predicate logic respectively to modal logic would collapse necessity into truth. The arguments are simple.

Suppose (2) would hold also for modal contexts. Then in particular $\models (A \Leftrightarrow B) \rightarrow (\Box A \Leftrightarrow \Box B)$. Taking for A the expression $a = a$, it follows that $\models B \rightarrow \Box B$, since both $a = a$ and $\Box(a = a)$ hold. Because we usually also assume the converse, $\models \Box B \rightarrow B$, it follows that $\models B \Leftrightarrow \Box B$.

Suppose (3) would hold also for modal contexts. Next suppose B is true. Taking $P(a) := a = a$ and $Q(a) := a = a \wedge B$, $P(a) \Leftrightarrow Q(a)$ is true and hence, by principle (3), $\Box(a = a) \Leftrightarrow \Box(a = a \wedge B)$ is true. Since $\Box(a = a \wedge B)$ is equivalent to $\Box(a = a) \wedge \Box B$, it follows that $\Box B$ is true. So, we have shown that from principle (3) it follows that $\models B \rightarrow \Box B$ and therefore $\models B \Leftrightarrow \Box B$.

Consequently, the principles (2) and (3) should NOT hold for modal contexts. In other words, modal contexts should be *extensionally opaque*; that is, – formulated negatively – general terms and sentences with the same extension (truth value in the case of sentences) must in general not be interchangeable with preservation of truth. Such interchangeability would amount to the collapse of modal distinctions. Formulated positively, extensional opacity means that some properties belong to things necessarily, while other properties belong to things only accidentally.

So, in order to make sense of quantified modal logic, modal contexts should be *referentially transparent*, i.e., principle (1) should hold also for modal contexts, and at the same time they should be *extensionally opaque*, i.e., the principles (2) and (3) should NOT hold for modal contexts. From this it is immediately clear that a satisfactory semantics for the modalities must distinguish between expressions which refer (singular terms) and expressions which have extension (general terms and sentences, the extension of a sentence being its truth value). Therefore, a Fregean semantics, according to which all expressions are considered to be referring, cannot be appropriate for modal logic. However, as already has been pointed out by J.R. Searle, Frege's extension of the notion of reference to predicates and sentences is not very natural:

... an expression refers to an object only because it conveys something true of that object.
But a predicate does not convey something true of a concept nor does a sentence convey something true of a truth value. [Searle [34], p. 3]

Summarizing, if we want quantified modal logic to make sense, we have to accept principle (1) also for modal contexts; in other words, modal contexts should be *referentially transparent*: whatever is true of an object is true of it regardless of how it is referred to (α). On the other hand, in order to avoid that necessity collapses into truth, we should not accept the principles (2) and (3) for modal contexts. In other words, modal contexts should be *extensionally opaque*: among the predicates true of an object, some are necessarily true of it, others only accidentally (β). And *essentialism* is just this combination of (α) and (β). See also Perrick [31].

6.12 The Modal Logic *GL*

The axioms of the modal logic *GL* (Gödel's Logic, also called the *Logic of Provability*) are the following:

the axioms of classical propositional logic;

$$\Box(B \rightarrow C) \rightarrow (\Box B \rightarrow \Box C);$$

$$\Box A \rightarrow \Box(\Box A); \text{ and}$$

$$\Box(\Box A \rightarrow A) \rightarrow \Box A.$$

The two rules of GL are Modus Ponens and necessitation (from $\vdash A$ infer $\vdash \Box A$).

The axioms and rules of GL resemble facts about $Prov(a)$, in particular (i), (ii), (iii) and (iv) in Subsection 5.2.2:

$$(i) \text{ if } \mathcal{P} \vdash A, \text{ then } \mathcal{P} \vdash Prov(\ulcorner A \urcorner);$$

$$(ii) \mathcal{P} \vdash Prov(\ulcorner B \rightarrow C \urcorner) \rightarrow (Prov(\ulcorner B \urcorner) \rightarrow Prov(\ulcorner C \urcorner));$$

$$(iii) \mathcal{P} \vdash Prov(\ulcorner A \urcorner) \rightarrow Prov(\ulcorner Prov(\ulcorner A \urcorner) \urcorner).$$

$$(iv) \text{ if } \mathcal{P} \vdash Prov(\ulcorner A \urcorner), \text{ then } \mathcal{P} \vdash A.$$

However, $Prov(a)$ does NOT meet the stronger condition $\mathcal{P} \vdash Prov(\ulcorner A \urcorner) \rightarrow A$.

Theorem 6.10. $\vdash A$ in GL iff for every Kripke model $M = \langle W, R, \models \rangle$ with W finite and non-empty, R transitive and irreflexive (i.e., for all $w \in W$, not wRw), $M \models A$.

For a proof of this theorem the reader is referred to Boolos, Burgess and Jeffrey, [5], Chapter 27. Here we restrict ourselves to the remark that $\Box(B \rightarrow C) \rightarrow (\Box B \rightarrow \Box C)$ holds in any Kripke model and that $\Box A \rightarrow \Box \Box A$ holds in any Kripke model $M = \langle W, R, \models \rangle$ with R transitive. In Exercise 6.30 the reader is asked to prove that $\Box(\Box A \rightarrow A) \rightarrow \Box A$ holds in any finite Kripke model $M = \langle W, R, \models \rangle$ with R transitive and irreflexive. Note that $\Box A \rightarrow A$ does NOT hold in such Kripke models, which corresponds to the fact that NOT $\mathcal{P} \vdash Prov(\ulcorner A \urcorner) \rightarrow A$. The weaker statement (iv) 'if $\mathcal{P} \vdash Prov(\ulcorner A \urcorner)$, then $\mathcal{P} \vdash A$ ' does hold, which corresponds to the fact that if $\vdash \Box A$ in GL , then also $\vdash A$ in GL .

Definition 6.22. Let ϕ be a function that assigns to each atomic formula of modal propositional logic a sentence in the formal language \mathcal{L}_A for arithmetic. For any formula A of modal propositional logic, the formula A^ϕ in \mathcal{L}_A is inductively defined as follows: $P_i^\phi := \phi(P_i)$ for any atomic formula P_i , $i = 1, 2, \dots$;

$$(B \rightarrow C)^\phi := B^\phi \rightarrow C^\phi \quad ;$$

$$\perp^\phi := \bar{0} = \bar{1} \quad ;$$

$$(\Box B)^\phi := Prov(\ulcorner B^\phi \urcorner) .$$

\wedge , \vee and \neg are treated similarly to \rightarrow .

The following theorems bring out an important connection between the formal system GL of modal logic and the formal system \mathcal{P} for arithmetic.

Theorem 6.11 (Arithmetical Soundness).

$$\text{If } \vdash A \text{ in } GL, \text{ then for all } \phi, \mathcal{P} \vdash A^\phi.$$

Proof. We restrict ourselves to the following observations:

If A is an axiom of propositional logic, then clearly $\mathcal{P} \vdash A^\phi$.

Let A be $\Box(B \rightarrow C) \rightarrow (\Box B \rightarrow \Box C)$. Then, by (ii) above, $\mathcal{P} \vdash A^\phi$.

Let A be $\Box B \rightarrow \Box \Box B$. Then, by (iii) above, $\mathcal{P} \vdash A^\phi$.

Corresponding to Modus Ponens: if $\mathcal{P} \vdash A^\phi$ and $\mathcal{P} \vdash (A \rightarrow B)^\phi$, then $\mathcal{P} \vdash B^\phi$.

Corresponding to the necessitation rule of GL : if $\mathcal{P} \vdash A^\phi$, then, by (i) above, also

$\mathcal{P} \vdash (\Box A)^\phi$, i.e., $\mathcal{P} \vdash \text{Prov}(\ulcorner A^\phi \urcorner)$.

It thus remains to show that $\mathcal{P} \vdash A^\phi$, where A is an axiom $\Box(\Box B \rightarrow B) \rightarrow \Box B$. For a proof of this the reader is referred to Boolos, Burgess and Jeffrey [5], Chapter 27.

Theorem 6.12 (Arithmetical completeness theorem).

If for all ϕ , $\mathcal{P} \vdash A^\phi$, then $\vdash A$ in GL .

This was proved by R. Solovay [35] and is also proved in Boolos [6], Chapter 12.

Exercise 6.30. Show that $\Box(\Box A \rightarrow A) \rightarrow \Box A$ holds in any Kripke model $M = \langle W, R, \models \rangle$ with W finite and non-empty, R transitive and irreflexive.

6.13 Solutions

Solution 6.1. Depending on what the speaker has in mind, at least two translations are possible: $\Box(S \rightarrow E), \Box(E \rightarrow L), \neg\Diamond L \vdash^? \neg\Diamond S$, and $S \rightarrow \Box E, E \rightarrow \Box L, \neg\Diamond L \vdash^? \neg\Diamond S$. The first argument is correct, the second incorrect; see Exercise 6.19.

Solution 6.2. a) In Example 6.1 we have seen that $\vdash A \rightarrow \Diamond A$ in KT and $\Diamond A \rightarrow \Box\Diamond A$ is an axiom of $S5$. By propositional logic: $A \rightarrow \Diamond A, \Diamond A \rightarrow \Box\Diamond A \vdash A \rightarrow \Box\Diamond A$. Therefore, $\vdash A \rightarrow \Box\Diamond A$ in $S5$. b) $\Diamond\neg A \rightarrow \Box\Diamond\neg A$ is an axiom of $S5$ and by propositional logic $\vdash \Diamond\neg A \Leftrightarrow \neg\Box A$. Hence, $\vdash \neg\Box A \rightarrow \Box\neg\Box A$ in $S5$. This is called *negative introspection*: if I do not know A , then I know that I do not know A .

Solution 6.3. $\Diamond P, \Box(\Diamond P \rightarrow Q) \vdash \Box Q$ in $S5$:

prem	axiom $S5$	prem	axiom
$\Diamond P$	$\Diamond P \rightarrow \Box\Diamond P$	$\Box(\Diamond P \rightarrow Q)$	$\Box(\Diamond P \rightarrow Q) \rightarrow (\Box\Diamond P \rightarrow \Box Q)$
$\Box\Diamond P$		$\Box\Diamond P \rightarrow \Box Q$	
$\Box Q$			

Solution 6.4.

By propositional logic, $(Q \rightarrow \Box Q) \vdash (\neg\Box Q \rightarrow \neg Q)$.
 So, by the \Box -axiom and MP , $\Box(Q \rightarrow \Box Q) \vdash \Box(\neg\Box Q \rightarrow \neg Q)$ in K .
 Again by the \Box -axiom and MP , $\Box(Q \rightarrow \Box Q) \vdash (\Box\neg\Box Q \rightarrow \Box\neg Q)$ in K . (1)
 According to Exercise 6.2: $\vdash \neg\Box Q \rightarrow \Box\neg\Box Q$ in $S5$. But $\vdash \neg A \rightarrow B$ iff $\vdash A \vee B$ by propositional logic; therefore: $\vdash \Box Q \vee \Box\neg\Box Q$ in $S5$. (2)
 From (1) and (2): $\Box(Q \rightarrow \Box Q) \vdash (\Box Q \vee \Box\neg Q)$ in $S5$.
 Hence, by propositional logic, $\Box(Q \rightarrow \Box Q), \Diamond Q \vdash \Box Q$ in $S5$.

Solution 6.5. The mistake is made in the transition from 3. to 4.: from the premiss $Q \rightarrow \Box Q$ it follows that $\neg\Box Q \rightarrow \neg Q$; but we do not have $\vdash \neg\Box Q \rightarrow \neg Q$ and therefore we cannot apply the rule of necessitation, which would yield $\vdash \Box(\neg\Box Q \rightarrow \neg Q)$ and then, by the axiom for \Box and MP , $\vdash \Box\neg\Box Q \rightarrow \Box\neg Q$.

Solution 6.6. $A \rightarrow A \vee B$ is an axiom of propositional logic, so $\vdash A \rightarrow A \vee B$. Hence, by the \Box -rule, $\vdash \Box(A \rightarrow A \vee B)$ in K . Hence, by the \Box -axiom of K and Modus Ponens, $\vdash \Box A \rightarrow \Box(A \vee B)$ in K .

We want to show that $\vdash \Diamond A \rightarrow \Diamond(A \vee B)$ in K . By contraposition it suffices to show that $\vdash \Box\neg(A \vee B) \rightarrow \Box\neg A$. We know from propositional logic that $\vdash \neg(A \vee B) \rightarrow \neg A$. By the \Box -rule it follows: $\vdash \Box(\neg(A \vee B) \rightarrow \neg A)$ in K . Therefore, by the \Box -axiom and Modus Ponens, $\vdash \Box\neg(A \vee B) \rightarrow \Box\neg A$ in K .

Solution 6.7. Let $M = \langle \{w_0, w_1\}, R, \models \rangle$ be the Kripke model (for K) with $w_0 R w_1$, $w_0 \models P$ and not $w_1 \models P$. Then $M, w_0 \models P$, but not $M, w_0 \models \Box P$.

Solution 6.8. a) To show: for all Kripke models M and for all w in M , $M, w \models \Box(A \wedge B)$ iff $M, w \models \Box A \wedge \Box B$. This is true because for any w' in M with $w R w'$, $M, w' \models A \wedge B$ iff $M, w' \models A$ and $M, w' \models B$.

b) However, $\not\models \Box(A \vee B) \Leftrightarrow (\Box A \vee \Box B)$. $M = \langle \{w_0, w_1\}, R, \models \rangle$ with R reflexive, $w_0 R w_1$, $w_0 \models P$ and $w_1 \models Q$, is a counterexample: $M, w_0 \models \Box(P \vee Q)$, but $M, w_0 \not\models \Box P$ and $M, w_0 \not\models \Box Q$.

c) To show: for all Kripke models M and for all w in M , $M, w \models \Diamond(A \vee B)$ iff $M, w \models \Diamond A$ or $M, w \models \Diamond B$. This is true because: there is a w' with $w R w'$ such that $M, w' \models A \vee B$ iff there is w' with $w R w'$ such that $M, w' \models A$ or there is a w' with $w R w'$ such that $M, w' \models B$.

Solution 6.9. By definition, $M, w_1 \models Q$.

$M, w_1 \models K_A P$, because w_1, w_2, w_4 are accessible from w_1 for Alice and $M, w_1 \models P$, $M, w_2 \models P$ and $M, w_4 \models P$.

$M, w_1 \models \neg K_B P$, because $w_1 R_B w_3$ and $M, w_3 \not\models P$.

$M, w_1 \models \neg K_A Q$ and $M, w_1 \models \neg K_B Q$ are shown in a similar way.

$M, w_1 \models K_A K_A P := M, w_1 \models K_A P$ and $M, w_2 \models K_A P$ and $M, w_4 \models K_A P$, which are all true.

$M, w_1 \models K_B K_A P := M, w_1 \models K_A P$ and $M, w_3 \models K_A P$ and $M, w_4 \models K_A P$. Because $w_3 R_A w_3$ and $M, w_3 \not\models P$, it follows that $M, w_3 \not\models K_A P$ and hence $M, w_1 \not\models K_B K_A P$, i.e., $M, w_1 \models \neg K_B K_A P$.

$M, w_1 \models K_A \neg K_B P := M, w_1 \models \neg K_B P$ and $M, w_2 \models \neg K_B P$ and $M, w_4 \models \neg K_B P$, because w_1, w_2, w_4 are accessible from w_1 for Alice. However, $M, w_2 \models K_B P$ and $M, w_4 \models K_B P$. Hence, $M, w_1 \not\models K_A \neg K_B P$, i.e., $M, w_1 \models \neg K_A \neg K_B P$.

The other cases are treated similarly.

Solution 6.10. $\neg(W \rightarrow \Box S) \vdash' \Diamond(W \rightarrow \neg S)$ in K , since the following tableau in K is closed:

$T \neg(W \rightarrow \Box S), F \Diamond(W \rightarrow \neg S)$

$F(W \rightarrow \Box S), T \Box \neg(W \rightarrow \neg S)$

$TW, F \Box S, T \Box \neg(W \rightarrow \neg S)$

$\overline{FS, T \neg(W \rightarrow \neg S)}$

$FS, F(W \rightarrow \neg S)$

$FS, TW, F \neg S$

FS, TW, TS

Solution 6.11. $\diamond(J \rightarrow P) \not\vdash' J \rightarrow \diamond P$ in K , since we can construct a counterexample:

$$\begin{array}{l}
 T \neg \Box \neg (J \rightarrow P), F (J \rightarrow \diamond P) \\
 F \Box \neg (J \rightarrow P), TJ, F \neg \Box \neg P \quad w_1 J \\
 \hline
 F \Box \neg (J \rightarrow P), TJ, T \Box \neg P \quad | \\
 \hline
 F \neg (J \rightarrow P), T \neg P \quad \downarrow \\
 T(J \rightarrow P), FP \quad w_2 \\
 \hline
 FJ, FP
 \end{array}$$

$M = \langle \{w_1, w_2\}, R, \models \rangle$, with $w_1 R w_2$ and $w_1 \models J$, is a counterexample in K :
 $M, w_1 \models \diamond(J \rightarrow P)$, since $M, w_2 \models J \rightarrow P$, and $M, w_1 \models J$, but $M, w_1 \not\models \diamond P$.

Solution 6.12. The tableau in i) is a tableau-deduction of $K_i B$ from $K_i(A \vee B)$ and $K_i \neg A$ in K . The search tree in ii) yields a counterexample M in $S4$ against the conjecture $A \vee B, K_i \neg A \vdash' K_i B$ in $S4$.

$$\begin{array}{ll}
 \text{i) } \frac{T K_i(A \vee B), T K_i \neg A, F K_i B}{T(A \vee B), T \neg A, FB} & \text{ii) } \frac{T(A \vee B), T K_i \neg A, F K_i B}{TB, T K_i \neg A, T \neg A, F K_i B} \quad w_0 B \\
 TA, FA, FB \mid TB, FA, FB & \frac{TB, T K_i \neg A, FA, F K_i B}{T K_i \neg A, FB} \quad | \\
 & T K_i \neg A, T \neg A, FB \quad \downarrow \\
 & T K_i \neg A, FA, FB \quad w_1
 \end{array}$$

$M = \langle \{w_0, w_1\}, R, \models \rangle$, with $w_0 R w_1$ and $w_0 \models B$, is a countermodel in $S4$:
 $M, w_0 \models A \vee B$, $M, w_0 \models K_i \neg A$, but $M, w_0 \not\models K_i B$, since $M, w_1 \not\models B$.

Solution 6.13. $\Box A \vee \neg \Box A$ is tableau-provable in K , but $\Box A \vee \Box \neg A$ is not.

$\diamond A \vee \neg \diamond A$ is tableau-provable in KT , and $\diamond A \vee \diamond \neg A$ too:

$$\begin{array}{llll}
 \text{a) } \frac{F \Box A \vee \neg \Box A}{F \Box A, F \neg \Box A} & \text{b) } \frac{F \Box A \vee \Box \neg A}{F \Box A, F \Box \neg A} & \text{c) } \frac{F \diamond A \vee \neg \diamond A}{F \neg \Box \neg A, F \neg \neg \Box \neg A} & \text{d) } \frac{F \diamond A \vee \diamond \neg A}{F \neg \Box \neg A, F \neg \Box A} \\
 \hline
 FA, TA & \swarrow \quad \searrow & \frac{T \Box \neg A, F \Box \neg A}{T \neg A, F \neg A} & T \Box \neg A, T \Box A \\
 \text{closure} & & FA, TA & FA, TA
 \end{array}$$

The tableaux in a), c) and d) are closed, while the tableau in b) yields a Kripke counterexample $M = \langle \{w_0, w_1, w_2\}, R, \models \rangle$ in K with $w_0 R w_1$, $w_0 R w_2$, $w_1 \not\models A$ and $w_2 \models A$: $M, w_0 \not\models \Box A \vee \Box \neg A$.

Solution 6.14. Both tableaux below are closed and hence are a tableau-deduction of $\Box A \rightarrow \Box(A \vee B)$ and $\diamond A \rightarrow \diamond(A \vee B)$ in K , respectively.

$$\begin{array}{ll}
 \frac{F \Box A \rightarrow \Box(A \vee B)}{T \Box A, F \Box(A \vee B)} & \frac{F \diamond A \rightarrow \diamond(A \vee B)}{T \neg \Box \neg A, F \neg \Box \neg(A \vee B)} \\
 \hline
 TA, FA, FB & \frac{F \Box \neg A, T \Box \neg(A \vee B)}{F \neg A, T \neg(A \vee B)} \\
 \text{closure} & TA, FA \vee B \\
 & TA, FA, FB
 \end{array}$$

Solution 6.15. Suppose $\vdash' \Box A \vee \Box B$ in K , KT or $S4$, i.e., there is a closed tableau starting with: $F \Box A \vee \Box B$

$$F \Box A, F \Box B$$

This tableau will continue with either FA or FB and one of these two will be closed.

In the first case a tableau starting with $F \Box A$ will give closure and in the second case, a tableau starting with $F \Box B$ will give closure.

Solution 6.16.

a) tableau in KT for:	b) search tree in KT for:	countermodel M
$F \Diamond P \rightarrow \Diamond \Diamond P$	$F \Diamond \Diamond P \rightarrow \Diamond P$	
$T \neg \Box \neg P, F \neg \Box \Box \neg P$	$T \neg \Box \Box \neg P, F \neg \Box \neg P$	
$F \Box \neg P, T \Box \Box \neg P, \dots$	$F \Box \Box \neg P, T \Box \neg P, \dots$	w_0
$F \neg P, T \Box \neg P$	$F \Box \neg P, T \neg P$	\downarrow
$F \neg P, T \Box \neg P, T \neg P$	$F \Box \neg P, FP$	w_1
$TP, T \Box \neg P, FP$	$F \neg P$	\downarrow
closure	TP no closure	$w_2 P$

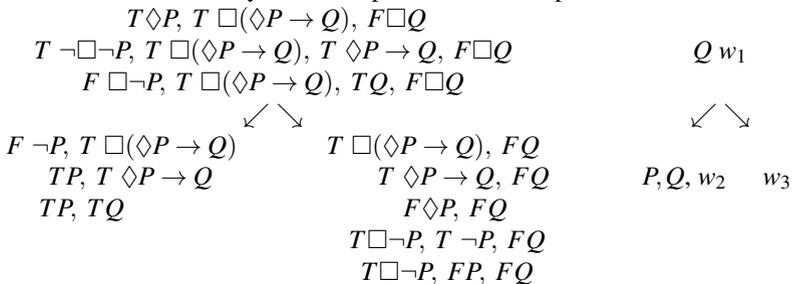
The tableau in a) closes and hence is a tableau-proof of $\Diamond P \rightarrow \Diamond \Diamond P$ in KT . $M = \langle \{w_0, w_1, w_2\}, R, \models \rangle$, with R reflexive, not transitive, $w_0 R w_1, w_1 R w_2$, not $w_0 R w_2$ and $w_2 \models P$, is a counterexample in KT against $\Diamond \Diamond P \rightarrow \Diamond P$: $M, w_0 \models \Diamond \Diamond P$, because $M, w_1 \models \Diamond P$, since $M, w_2 \models P$. But $M, w_0 \not\models \Diamond P$, because $M, w_1 \not\models P$.

c) search tree in $S4$ for:	countermodel M'	d) search tree in $S4$ for:	M''
$F P \rightarrow \Box \Diamond P$	$w_0 P$	$T P \rightarrow Q, F \neg \neg \Box \neg (P \wedge \neg Q)$	w_0
$TP, F \Box \neg \Box \neg P$	\downarrow	$FP, F \Box \neg (P \wedge \neg Q)$	\downarrow
$F \neg \Box \neg P$	w_1	$F \neg (P \wedge \neg Q)$	$w_1 P$
$T \Box \neg P, T \neg P$		$T(P \wedge \neg Q)$	
$T \Box \neg P, FP$		$TP, T \neg Q, FQ$	

$M' = \langle \{w_0, w_1\}, R, \models \rangle$, with $w_0 R w_1$ and $w_0 \models P$, is a countermodel in $S4$ for $P \rightarrow \Box \Diamond P$, because $M', w_0 \models P$, but $M', w_0 \not\models \Box \Diamond P$, since $M', w_1 \not\models \Diamond P$.

$M'' = \langle \{w_0, w_1\}, R, \models \rangle$, with $w_0 R w_1$ and $w_1 \models P$, is a countermodel in $S4$ for $(P \rightarrow Q) \rightarrow \neg \Diamond (P \wedge \neg Q)$, because $M'', w_0 \models P \rightarrow Q$, but $M'', w_0 \models \Diamond (P \wedge \neg Q)$, since $M'', w_1 \models P \wedge \neg Q$.

Solution 6.17. The following search tree for the conjecture $\Diamond P, \Box(\Diamond P \rightarrow Q) \vdash \Box Q$ in $S4$ does not close and hence yields a Kripke counterexample in $S4$:



$M = \langle \{w_1, w_2, w_3\}, R, \models \rangle$, with $w_1 R w_2, w_1 R w_3$, R reflexive and transitive, $w_1 \models Q$ and $w_2 \models P \wedge Q$, is a countermodel in $S4$ for the conjecture in question: $M, w_1 \models \Diamond P$, $M, w_1 \models \Box(\Diamond P \rightarrow Q)$ because in every world in which $\Diamond P$ is true, Q is true too, but $M, w_1 \not\models \Box Q$.

Solution 6.18. The following search tree for the conjecture $\Box(Q \rightarrow \Box Q), \Diamond Q \vdash \Box Q$ in $S4$ does not close and hence yields a counterexample in $S4$:

counterexample in $S4$ for both conjectures: $M, w_1 \models \neg A$, $M, w_1 \models B$, but $M, w_1 \not\models \Box(A \rightarrow B)$, since $M, w_2 \models A$ and $M, w_2 \not\models B$.

Solution 6.21. The following tableaux in K are all closed and hence:

$$\frac{\Box \neg A \vdash \Box(A \rightarrow B), \Box B \vdash \Box(A \rightarrow B), Q \vdash \Box(P \rightarrow P) \text{ and } \vdash \Box(\neg Q \wedge Q \rightarrow P) \text{ in } K:}{\frac{T \Box \neg A, F \Box(A \rightarrow B)}{T \neg A, F A \rightarrow B} \quad \frac{T \Box B, F \Box(A \rightarrow B)}{TB, F A \rightarrow B} \quad \frac{TQ, F \Box(P \rightarrow P)}{FP \rightarrow P} \quad \frac{F \Box(\neg Q \wedge Q \rightarrow P)}{F \neg Q \wedge Q \rightarrow P}}$$

$$\frac{}{FA, TA, FB} \quad \frac{}{TB, TA, FB} \quad \frac{}{TP, FP} \quad \frac{}{T \neg Q \wedge Q, FP}$$

$$\frac{}{FQ, TQ, FP}$$

Solution 6.22. a) $M, w \models C \Box \rightarrow (B_F \wedge V_F) \vee (B_I \wedge V_I)$: there is a world, namely w_1 (or w_2), such that $M, w_1 \models C \wedge ((B_F \wedge V_F) \vee (B_I \wedge V_I))$ and such that for any w'' , if $M, w'' \models C \wedge \neg((B_F \wedge V_F) \vee (B_I \wedge V_I))$, then $w_1 <_w w''$.

b) $M, w \not\models C \Box \rightarrow B_I$ because there is no $C \wedge B_I$ -world which is closer to w than any $C \wedge \neg B_I$ -world; w_1 is as close to w as w_2 .

c) and d) are treated similarly.

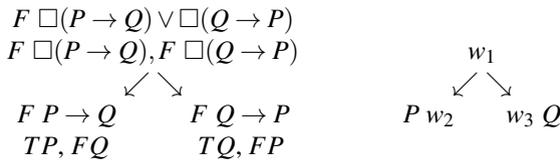
Solution 6.23. a) Suppose $M, w \models A \wedge \neg B$ (1) and $M, w \models A \Box \rightarrow B$. From (1) $M, w \models A$ and hence, since R is reflexive, $M, w \models \Diamond A$. So, from the definition of $M, w \models A \Box \rightarrow B$, it follows that there is some world w' in W such that (2) wRw' and $M, w' \models A \wedge B$, and (3) for all w'' in W , if $M, w'' \models A \wedge \neg B$, then $w' <_w w''$.

From (1) and (3) it follows that $w' <_w w$. And since $M, w \models \neg B$ and $M, w' \models B$ we know that $w \neq w'$ and therefore, by assumption, *not* $w' <_w w$. Contradiction. So, if $M, w \models A \wedge \neg B$, then $M, w \models \neg(A \Box \rightarrow B)$.

b) Suppose $M, w \models A \wedge B$. Since R is reflexive we have that wRw and $M, w \models A \wedge B$. So, in order to show that $M, w \models A \Box \rightarrow B$ it suffices to show that for all w'' in W , if $M, w'' \models A \wedge \neg B$, then $w <_w w''$. So, suppose $M, w'' \models A \wedge \neg B$. Now, $M, w \models B$ and $M, w'' \models \neg B$. Therefore, $w \neq w''$ and hence, by assumption, $w <_w w''$.

Solution 6.24. Suppose $M, w \models \Box(A \rightarrow B)$. If $M, w \models \neg \Diamond A$, then $M, w \models A \Box \rightarrow B$. So, suppose $M, w \models \Diamond A$, i.e., for some w' in W , wRw' and $M, w' \models A$. Since $M, w \models \Box(A \rightarrow B)$, it follows that wRw' and $M, w' \models A \wedge B$ (1). So, in order to show that $M, w \models A \Box \rightarrow B$ it suffices to prove that for all w'' in W , if $M, w'' \models A \wedge \neg B$, then $w' <_w w''$. So, suppose $M, w'' \models A \wedge \neg B$. Since $M, w \models \Box(A \rightarrow B)$ it follows that *not* wRw'' . Then, by assumption, it follows from wRw' and *not* wRw'' that $w' <_w w''$.

Solution 6.25. The following search tree for the conjecture that $\vdash \Box(P \rightarrow Q) \vee \Box(Q \rightarrow P)$ in $S4$ does not close and hence yields a Kripke counterexample in $S4$:



From this open search tree we can read off a Kripke counterexample in $S4$ to $P \Box \rightarrow Q \vee Q \Box \rightarrow P$. Let $M = \langle \{w_1, w_2, w_3\}, R, \models \rangle$ with R reflexive and transitive, $w_1 R w_2$, $w_1 R w_3$, $w_2 \models P$ and $w_3 \models Q$. Then $M, w_1 \not\models \Box(P \rightarrow Q) \vee \Box(Q \rightarrow P)$.

It happens that also $M, w_1 \not\models P \Box \rightarrow Q$, since $M, w_1 \not\models \neg \Diamond P$ and for no w in $\{w_1, w_2, w_3\}$, $M, w \models P \wedge Q$. In a similar way one sees that $M, w_1 \not\models Q \Box \rightarrow P$.

Solution 6.26. Deduction Theorem: $A^1, \dots, A^n \vdash^* B$ iff $A^1, \dots, A^{n-1} \vdash^* A^n \Rightarrow B$.

Proof: From right to left is trivial. From left to right: Suppose $A^1, \dots, A^n \vdash^* B$. Replace in the given deduction of B from A^1, \dots, A^n each expression C_c with the natural number n occurring in the index c by the expression $(A^n \Rightarrow C)_{c-\{n\}}$, where the index $c - \{n\}$ results from c by leaving out n . The upper lines in the resulting schema may look as follows:

$$A^1_{\{1\}}, \dots, A^{n-1}_{\{n-1\}}, A^n \Rightarrow A^n, \text{ axiom.}$$

The bottom line in the resulting schema just contains $A^n \Rightarrow B_{\{1, \dots, n-1\}}$. Note that

$A^n \Rightarrow A^n$ is an axiom. For Modus Ponens, $\frac{D_d (D \Rightarrow E)_e}{E_{d \cup e}}$ there are four possibilities:

i) n occurs in d , but not in e . Then we get $\frac{A^n \Rightarrow D_{d-\{n\}} (D \Rightarrow E)_e}{(A^n \Rightarrow E)_{d \cup e - \{n\}}}$ and using axiom

2 this is a derived rule.

ii) n occurs in e , but not in d . Then we get $\frac{D_d (A^n \Rightarrow (D \Rightarrow E))_{e-\{n\}}}{(A^n \Rightarrow E)_{d \cup e - \{n\}}}$ and by using

axiom 4 this is a derived rule.

iii) n occurs both in d and in e . Then we get $\frac{(A^n \Rightarrow D)_{d-\{n\}} (A^n \Rightarrow (D \Rightarrow E))_{e-\{n\}}}{(A^n \Rightarrow E)_{d \cup e - \{n\}}}$

and by using axiom 3 this is a derived rule.

iv) In case n occurs neither in d nor in e , the application of Modus Ponens remains unchanged. So, the resulting schema can be extended - by using the axioms - to a deduction of $A^n \Rightarrow B$ from A^1, \dots, A^{n-1} .

Solution 6.27. a) One easily checks that the axioms for weak implication are valid (1). For instance, let $M = \langle S, \emptyset, \cup, \models \rangle$ be a model; then $M \models_{\emptyset} A \Rightarrow A$, i.e., for all a in S , if $M \models_a A$, then $M \models_a A$. And the rule Modus Ponens preserves validity (2), more precisely: if $M \models_{\emptyset} B$ and $M \models_{\emptyset} B \Rightarrow C$, then $M \models_{\emptyset} C$.

Now suppose $A_1, \dots, A_n \vdash^* B$. Then, by the deduction theorem, $\vdash^* A_1 \Rightarrow (\dots \Rightarrow (A_n \Rightarrow B) \dots)$, i.e., the latter formula can be obtained by a finite number of applications of Modus Ponens starting with the axioms for weak implication. So, by (1) and (2), for all models M , $M \models_{\emptyset} A_1 \Rightarrow (\dots \Rightarrow (A_n \Rightarrow B) \dots)$, i.e., $A_1, \dots, A_n \models^* B$.

ii) Let $M = \langle \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}, \emptyset, \cup, \models \rangle$ be defined by $\{1\} \models Q$, $\{2\} \models P$ and $\{1, 2\} \not\models P$. Then $M \not\models_{\{1\}} P \Rightarrow P$ and $M \not\models_{\emptyset} Q \Rightarrow (P \Rightarrow P)$.

Solution 6.28. 1. $(A \Rightarrow A) \Rightarrow (A \Rightarrow A)$, axiom 1 for weak implication.

2. $((A \Rightarrow A) \Rightarrow (A \Rightarrow A)) \Rightarrow (A \Rightarrow ((A \Rightarrow A) \Rightarrow A))$, axiom 4 for weak implication.

3. $A \Rightarrow ((A \Rightarrow A) \Rightarrow A)$, from 1 and 2 by MP.

Solution 6.29. $\Box \forall x[A(x)] \vdash^{\prime} \forall x[\Box A(x)]$:
$$\begin{array}{l} T \Box \forall x[A(x)], F \forall x[\Box A(x)] \\ T \Box \forall x[A(x)], F \Box A(a) \\ T \forall x[A(x)], F A(a) \\ T A(a), F A(a) \text{ closure} \end{array}$$

Solution 6.30. Let $M = \langle W, R, \models \rangle$ be a Kripke model with W finite and non-empty, R transitive and irreflexive. Suppose $M, w \models \Box(\Box A \rightarrow A)$, i.e., for all w' , if wRw'

and $M, w' \models \Box A$, then $M, w' \models A$. (1)

Next suppose that not $M, w \models \Box A$. Then there is $w_1 \in W$ such that wRw_1 and not $M, w_1 \models A$. From (1) it follows that not $M, w_1 \models \Box A$. Hence, there is $w_2 \in W$ such that w_1Rw_2 and not $M, w_2 \models A$. Because R is transitive, wRw_2 . So, by (1), not $M, w_2 \models \Box A$. Consequently, there is $w_3 \in W$ such that w_2Rw_3 and not $M, w_3 \models A$. And so on.

So, we find a sequence $w = w_0, w_1, w_2, \dots$ in W such that w_iRw_{i+1} and not $M, w_i \models A$. Because R is transitive and irreflexive it follows that $w_i \neq w_j$ for all i, j with $i \neq j$. So, W is infinite. Contradiction.

References

1. Ackermann, W., Begründung einer strengen Implikation. *Journal of Symbolic Logic*, 21, 113-128, 1956.
2. Anderson, A.R., and Belnap, N.D., *Entailment; the Logic of Relevance and Necessity*. Princeton University Press, 1975.
3. Benthem, J.F.A.K. van, *The Logic of Time*. Reidel, Dordrecht, 1983.
4. Blackburn, P., M. de Rijke and Y. Venema, *Modal Logic*. Cambridge Univ. Pr. 2001.
5. Boolos, G., J. Burgess and R. Jeffrey, *Computability and Logic*. Cambridge University Press, 5th edition, 2007.
6. Boolos, G., *The Unprovability of Consistency; an essay in modal logic*. Cambridge University Press, 1979.
7. Burgess, J.P., Logic and Time. *Journal of Symbolic Logic* 44, 566 - 582, 1979.
8. Church, A., The weak Theory of Implication. In: Menne-Wilhelmy-Angsil (eds.), *Kontrolliertes Denken, Untersuchungen zum Logikkalkül und der Logik der Einzelwissenschaften*. Kommissions-Verlag Karl Alber, München, pp. 22-37, 1951.
9. Donnellan, K., Reference and Definite Descriptions. *Philosophical Review* 75 (1966) 281-304.
10. Fagin, R., J.Y. Halpern, Y. Moses and M. Vardi, *Reasoning About Knowledge*. MIT Press, 2004.
11. Fitting, M., *Proof Methods for Modal and Intuitionistic Logics*. Reidel, 1983.
12. Føllesdal, D., Essentialism and Reference. In: Hahn, L.E., and P.A. Schilpp (eds.), *The Philosophy of W.V. Quine*. La Salle, Illinois, 1986.
13. Frege, G., *Begriffsschrift*. Halle, 1879.
14. Gent, I.P., A sequent system for Lewis's counterfactual logic VC. *Notre Dame Journal of Formal Logic*, 33 (1992) 369-382.
15. Goldblatt, R., *Logics of Time and Computation*. CSLI, Stanford, 1987.
16. Grice, P., *Studies in the Way of Words*. Harvard University Press, 1989.
17. Harper, W.L., R.C. Stalnaker, G. Pearce (eds.), *IFS; Conditionals, Belief, Decision, Chance and Time*. Reidel, Dordrecht, 1976.
18. Hilpinen, R. (ed.), *Deontic Logic*. Introductory and Systematic Readings. Reidel, 1971.
19. Hubbeling, H.G., *Language, Logic and Criterion*. Van Gorcum, Assen, 1971.
20. Kneale, W. and M., *The Development of Logic*. Clarendon Press, Oxford, 1962.
21. Kripke, S.A., Semantical Analysis of Modal Logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, Band 9 (1963), 67-96.
22. Kripke, S.A., *Naming and Necessity*. Basil Blackwell, Oxford, 1980.
23. Kripke, S.A., *Identity and Necessity*. In: Schwartz, S.P., *Naming, Necessity and Natural Kinds*. Cornell University Press, 1977-1979, 66-101.
24. Lemmon, E.J., *An introduction to Modal Logic*. Basil Blackwell, Oxford, 1977.
25. Lewis, C.I., *A Survey of Symbolic Logic*, University of California Press, Berkeley, 1918.
26. Lewis, D., *Counterfactuals*. Harvard University Press, 1973.

27. Lewis, D., Counterfactuals and Comparative Possibility. *Journal of Philosophical Logic*, pp. 418-446, 1973.
28. Linsky, L. (ed.), *Reference and Modality*. Oxford University Press, London, 1971, 1979.
29. Meyer, J.-J. Ch. and W. van der Hoek, *Epistemic Logic for A.I. and Computer Science*. Cambridge University Press, 2004.
30. Montague, R., The Proper Treatment of Quantification in Ordinary English. Reprinted in R.H. Thomason, (ed.), *Formal Philosophy; Selected Papers of Richard Montague*. Yale University Press, London, pp. 247-270, 1974.
31. Perrick, M. and H.C.M. de Swart, Quantified Modal Logic, Reference and Essentialism. *Logique et Analyse* 143-144 (1993) 219-231.
32. Quine, W.V., Reference and Modality, *Journal of Symbolic Logic* (1953), 137-138.
33. Quine, W.V., *From a Logical Point of View*. Harvard University Press, 1980.
34. Searle, J.R. (ed.), *The Philosophy of Language*. Oxford University Press, 1971.
35. Solovay, R.M., Provability Interpretations of Modal Logic. *Journal of Symbolic Logic* 46, 661-662, 1981.
36. Swart, H.C.M. de, Systems of Natural Deduction for several systems of Modal Logic; constructive completeness proofs and effective decision procedures. *Logique et Analyse* 90-91, 263-284, 1980.
37. Swart, H.C.M. de, A System of Natural Deduction, an effective decision procedure and a constructive completeness proof for the counterfactual logics VC and VCS. *Journal of Symbolic Logic* 48, 1-20, 1983.
38. Urquhart, A., Semantics for Relevant Logics. *Journal of Symbolic Logic*, 37, 159-169, 1972.