An Introduction to Modal Logic

1. Introduction

In first order logic a formula is either *true* or *false* in any model; no other possibilities are allowed. In natural language, we distinguish between various "modes" of truth, e.g, "known to be true", "believed to be true", "necessarily true", "true in the future". Some examples bellow show such situations:

- " Barack Obama is the president of the US" is currently true but it will not be true at some point in the future.
- "After program P is executed, A hold" is possibly true if the program performs what is intended to perform.

Classical logic is *truth-functional*, namely truth value of a larger formula is determined by the truth value(s) of its subformula(e) via truth tables for $\wedge, \vee$, $\neg$, and $\rightarrow$.

In the 1920s, C.A. Lewis (a philosopher) tried to capture *a non-truth-functional notion* of "A Necessarily Implies B"? ($A \rightarrow B$)

We can take $A \rightarrow B$ to mean "it is impossible for A to be true and B to be false"

He chose a symbol, P, and wrote PA for "A is possible"; then:

$\neg PA$ is "A is impossible"

$\neg P \neg A$ is "not-A is impossible"

Then he used the symbol N to stand for $\neg P$ and expressed

$NA := \neg P \neg A$ "A is necessary"

Because $\rightarrow$ is logical implication, then we can transform it like in the following:

$A \rightarrow B := N(A \rightarrow B) = \neg P \neg (A \rightarrow B) = \neg P \neg (\neg A \vee B) = \neg P(A \wedge \neg B)$

So Lewis achieved his aim.

The two symbols P and N can be viewed as "possibly true" and "necessarily true" and are modes of truth. They are the basis of modal logics.

In modal logics, these modes of truth are represented by two operators: □ ◊

To simplify writing, one can also use two adjacent symbols to represent each operator: [] and <>

There are several names of the modal operators:

- In basic modal logic: □ is called box and ◊ is called diamond
- In the logics which study the necessity / possibility □ is called necessary and ◊ is called possible
- In logics about knowledge □ may be seen as what an agent knows or, alternately, what an agent believes (what an agent thinks it is true).

Some other modal operators can be introduced to talk about time, as in temporal logic (the modal logic of time), or about actions, as in dynamic logic (the modal logic of action).

Propositional modal logic is classical propositional logic extended with modal operators. We shall see the exact syntax and semantics of the modal logic in a few moments. There is also predicate modal logic, which is far more complicated.

2. Why do we need modal logic in computer science?

Modal logic is often used to reason about "modes of truth", for example as in temporal logic where we can express computational behavior of programs by distinguishing between truth at various points of program execution in the future, or as in dynamic logic where we can use the modal operators to formally express program properties, such as the execution of a sequence will certainly make a variable true or will possible make a variable true (or equal to some given value). Another example is to use a logic of knowledge and belief to express truth in distributed systems, or different beliefs on the truth of propositions hold by different agents.

Modal logic is certainly a powerful tool to represent and reason about knowledge in AI. As predicate logic has more power to represent knowledge as propositional logic (we can not express the statement "all men are mortal" in PL), as nonmonotonic logics allow us to capture various forms of common sense reasoning (for example default reasoning) which can not be conveniently expressed in FOPL, modal logic allow us to reason about the truth of some statements and model complex problems.

Some examples of logical puzzles follow, to give a glimpse of what the power of modal logics may bring in representing a problem solving process. Try finding the answer of these puzzles by yourself.

The Lady and the Tiger Puzzle (from Raymond Smullyan book "The Lady and the Tiger", Penguin Books 1984).

There are two rooms, A and B, with the following signs on them:

A: In this room there is a lady, and in the other room there is a tiger"

B: "In one of these rooms there is a lady and in one of them there is a tiger"

One of the two signs is true and the other one is false.

Q: Behind which door is the lady?

The King's Wise Men Puzzle

The King called the three wisest men in the country to his court to decide who would become his new advisor.

He painted a spot on each of their foreheads and told them that at least one of them has a white spot on his forehead. Actually all three of them had white spots on their foreheads but each wise man could see all of the other foreheads, but none of them could see their own. The king declared that whichever man finds out that the color of his spot is white would become his new advisor.

The first wise man said: “I do not know whether I have a white spot”, and the second man then says “I also do not know whether I have a white spot”. The third man says then “I know I have a white spot on my forehead”.

Q: How did the third wise man reason?

There are other versions of this puzzle, for example:

The King's Wise Men Puzzle – Version 1

The wise men were forbidden to speak to each other. The king asked “Who has a white spot on his forehead?” No answer came. The king asked a second time: “Who has a white spot on his forehead?” and again, no answer was obtained. Finally, the king asked a third time: “Who has a white spot on his forehead?” and the wisest man answered “I know I have a white spot on my forehead”.

Q: How did the wise man reason in this case?

OR

A certain king wishes to determine which of his three wise men is the wisest. He arranges them in a circle so that they can see and hear each other and tells them that he will put a white or black spot on each of their foreheads but that at least one spot will be white. In fact all three spots are white. He then offers his favor to the one who will first tell him the color of his spot. After a while, the wisest announces that his spot his white. How does he know?

The intended solution is that the wisest reasons that if his spot were black, the second would see a black and a white and would reason that if his spot were black, the third would have seen two black spots and reasoned from the king's announcement that his spot was white. This traditional version requires the wise men to reason about how fast their colleagues’ reason, and we don't wish to try to formalize this.

The Hat Puzzle

A team of N players, where N is at least three, are randomly assigned hats that are equally likely to be white or black, under conditions such that:

There is at least one black and one white hat assigned;

Each team player can see the hats of the other team members, but cannot see their own hat;

Team members cannot communicate in any way.

Hypothetically, there are N prisoners, but there was not enough space for all of them. The jailer decides to give them a test, and if all of them succeed in answering it, he will release them, whereas if any one of them answers incorrectly, then he will kill all of them. He describes the test as follows:

I will put a hat, either white or black, on the head of each of you. You can see others' hats, but you can't see your own hat. You are given 20 minutes. I will place at least one white hat and at least one black hat. All of you should tell me the color of the hat on your head. You can't signal to others or give a hint or anything like that. You should say only WHITE or BLACK. You can go and discuss for a while now.

All of them go and discuss for some time. And after they come back, he starts the test.

Interestingly, each of them answers correctly and hence all are released.

Q: How did they reason about it?

See more about this puzzle at http://en.wikipedia.org/wiki/Hat_Puzzle

Mr. S. and Mr. P Puzzle (Hans Freudenthal, Nieuw Archief Voor Wiskunde, Series 3, Volume 17, 1969, page 152)

Two numbers m and n are chosen such that $2 \leq m \leq n \leq 99$. Mr. S is told their sum and Mr. P is told their product. The following dialogue ensues:

Mr. P: I don't know the numbers.

Mr. S: I knew you didn't know. I don't know either.

Mr. P: Now I know the numbers.

Mr S: Now I know them too.

Q: In view of the above dialogue, what are the numbers?

Comments on **Mr. S. and Mr. P Puzzle**

By Martin Gardner, “A Pride of Problems, Including One that is Virtually Impossible”,

Scientific American, Volume 241, December 1979:

This beautiful problem, which I call "impossible" because it seems to lack sufficient information for a solution, began making the rounds of mathematics meetings a year or so ago. I do not know its origin. Mel Stover of Winnipeg was the first to call it to my attention.

Two numbers (not necessarily different) are chosen from the range of positive integers greater than 1 and not greater than 20. Only the sum of the two numbers is given to mathematician S. Only the product of the two is given to mathematician P.

On the telephone S says to P: "I see no way you can determine my sum."

An hour later, P calls back to say: "I know your sum."

Later S calls P again to report: "Now I know your product."

What are the two numbers?

To simplify the problem, I have given it here with an upper bound of 20 for each of the two numbers. This means that the sum cannot be greater than 40 or the product greater than 400. If you succeed in finding the unique solution, you will see how easily the problem can be extended by raising the upper bound. Surprisingly, if the bound is raised to 100, the answer remains the same. Stover tells me that a computer program in Israel checked on all numbers up to two million without finding a second solution. It may be possible to prove that the solution is unique even if there is no upper bound whatsoever.

The logical formalization of the King's Wise Men and Mr. S. and Mr. P puzzles can be found at Two Puzzles Involving Knowledge, McCarthy, John (1987).

http://www-formal.stanford.edu/jmc/puzzles.html

3. Syntax of Modal Logic

Atomic Formulae: p ::= p0 | p1 | p2 | …. where p_i are atoms in PL

Formulae: $\varphi ::= p \mid \neg\varphi \mid \Box\,\varphi \mid \Diamond\,\varphi \mid \varphi \wedge \psi \mid \varphi \vee \psi \mid \varphi \rightarrow \psi$ where φ and ψ are wffs in PL

Examples: $\Box p_0 \rightarrow p_2$ $\Box p_3 \rightarrow \Box\Box p_1$ $\Box\,(p_1 \rightarrow p_2) \rightarrow ((\Box p_1) \rightarrow (\Box p_2))$

Variables: p, q, r stand for atomic formulae while φ, ψ possibly with subscripts stand for arbitrary formulae (including atomic ones)

4. Deductions in Modal logics

L_M – the language of modal logic (wffs)

Axioms

The 3 axioms of PL

A1. $\varphi \rightarrow (\psi \rightarrow \varphi)$
A2. $(\varphi \rightarrow (\psi \rightarrow \zeta)) \rightarrow ((\varphi \rightarrow \psi) \rightarrow (\varphi \rightarrow \zeta))$
A3. $((\neg\varphi) \rightarrow (\neg\psi)) \rightarrow (\psi \rightarrow \varphi)$

The axiom to specify distribution of necessity

A4. $\Box\ (\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$ ***Axiom K (distribution)***

Inference rules

- Substitution (uniform) $\langle\varphi\rangle \xrightarrow{R} \varphi'$

Substitution is applied to schema of formulae.

Schemas: $\Box\varphi \rightarrow \varphi$ $\Box\varphi \rightarrow \Box\Box\,\varphi$ $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$

Schema Instances: Uniformly replace/substitute the formula variables with formulae

Examples: $\Box p_0 \rightarrow p_0$ is an instance of $\Box\varphi \rightarrow \varphi$ but $\Box p_0 \rightarrow p_2$ is not

- Modus Ponens : : $\langle\varphi, (\varphi \rightarrow \psi)\rangle \xrightarrow{R} \psi$

- The modal rule of necessity $<\varphi> \xrightarrow{R} \Box\varphi$ « for any formula φ, if φ was proved then you can infer $\Box\varphi$ » (***Necessity rule***)

Some other axioms may be added (will comment upon them later):

(D) $\Box\ \varphi \to \Diamond\varphi$
(T) $\Box\ \varphi \to \varphi$
(4) $\Box\ \varphi \to \Box\ \Box\varphi$
(B) $\varphi \to \Box\ \Diamond\ \varphi$
(5) $\Diamond\ \varphi \to \Box\ \Diamond\ \varphi$

System name	**Representation**
K	K
D	KD
T	KT
S4	KT4
B	KTB
S5	KT45

5. Semantics of Modal Logic

The semantics of modal logic is defined based on a *nonlinear model*, as opposed to the semantics of PL or FOPL, which is based on a linear model and interpretations on that model. The semantics of modal logic is known as the *Kripke Semantics*, also called the Possible World approach.

The semantics give an intuitive meaning to syntactic symbols and, more precisely, formally defines what it means for a formula to be true in this logic. Moreover, it allows defining the meaning of logical consequence in modal logic and thus a way to find out how we can derive new knowledge based on existing one.

Kripke semantics can be easily understood and defined using notions of graph theory. Remember that a *directed graph* (V, E) is defined as a set of vertices $V = \{v, v_1, v_2, \ldots\}$ and a set of (directed) edges $\{(s_1,t_1), (s_2,t_2),\ldots\}$ from the source vertex $s_i \in V$ to the target vertex $t_i \in V$ for $i = 1,2,\ldots$

If we consider the *cross product* of a set V, V x V stands for $\{(v,w) \mid v\in V \text{ and } w\in V\}$ the set of all ordered pairs (v,w), where v and w are from V. Using the notion of cross product we can define a directed graph as being a pair (V,E), where $V = \{v, v_1, v_2, \ldots\}$ and $E \subseteq V \times V$ is a binary relation over V.

A ***Kipke frame*** is a directed graph <W, R>, where W is a non-empty set of worlds (points, vertices) and $R \subseteq W \times W$ is a binary relation over W, called the *accessibility relation*.

An *interpretation* of a wff in modal logic on a Kripke frame <W, R> is a function $I : W \times L \to \{t,f\}$ which tells the truth value of every atomic formula from the language L at every point (in every word) in W.

A ***Kripke model M of a formula*** φ (an interpretation which makes the formula true) is the triple <W, R, I>, where I is an interpretation of the formula on a Kripke frame <W,R> which makes the formula true. This is denoted by $M \models_W \varphi$

Using the model, we can define the semantics of formulae in modal logic and can compute the truth value of formulae.

$M \models_W \neg\varphi$ iff $M \not\models_W \varphi$ (or $M \models_W \neg\varphi$)

$M \models_W \varphi \wedge \psi$ iff $M \models_W \varphi$ *and* $M \models_W \psi$

$M \models_W \varphi \vee \psi$ iff $M \models_W \varphi$ *or* $M \models_W \psi$

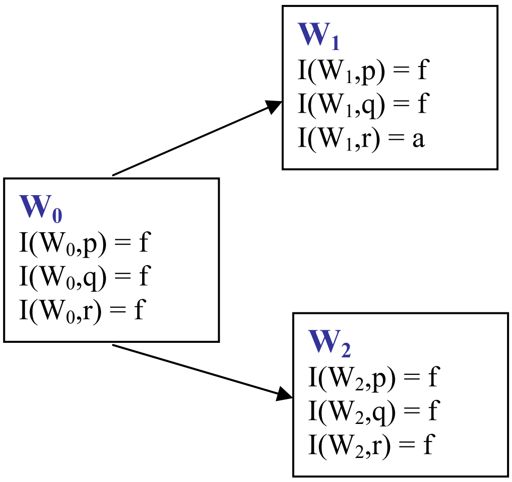
$M \models_W \varphi \rightarrow \psi$ iff $M \models_W \neg\varphi$ *or* $M \models_W \psi$ ($\neg\varphi \vee \psi$ is true in W)

$M \models_W \Diamond \varphi$ iff $\exists w'$: $R(w,w')$ & $M \models_{W'} \varphi$

$M \models_W \Box \varphi$ iff $\forall w'$: $R(w,w') \Rightarrow M \models_{W'} \varphi$

Example 1

p – I am rich
q – I am president of Romania
r – I am holding a PhD in CS



$I(W_0, \Diamond p) = f$ $I(W_0, \Box p) = f$

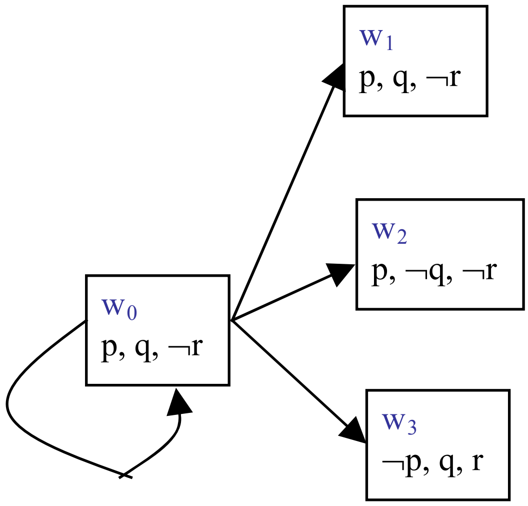
$I(W_0, \Diamond r) = t$ $I(W_0, \Box r) = f$

Example 2

p -Alice visits Paris

q - It is spring time

r - Alice is in Italy



$I(W_0, \Diamond p) = ?$	$I(W_0, \Box p) = ?$
$I(W_0, \Diamond q) = ?$	$I(W_0, \Box q) = ?$
$I(W_0, \Diamond r) = ?$	$I(W_0, \Box r) = ?$
$I(W_1, \Diamond p) = ?$	$I(W_1, \Box p) = ?$
$I(W_1, \Diamond q) = ?$	$I(W_1, \Box q) = ?$
$I(W_1, \Diamond r) = ?$	$I(W_1, \Box r) = ?$

Properties of formulae

Let M= <W,R,I> be any Kripke model, φ a formula and $w \in W$.

Example: If $I(w, \Box\varphi) = t$ then $I(w, \Diamond\neg\varphi) = f$

Example: If $I(w, \Diamond \neg \varphi) = f$ then $I(w, \neg \Diamond \neg \varphi) = t$ $\qquad \Box\varphi \rightarrow \neg\Diamond\neg\varphi$

Example: If $I(w, \Diamond\varphi) = t$ then $I(w, \Box \neg\varphi) = f$

Example: If $I(w, \Box \neg\varphi) = f$ then $I(w, \neg \Box \neg\varphi) = t$ $\qquad \Diamond\varphi \rightarrow \neg\Box\neg\varphi$

Exercise: Show that all these implications are reversible.

Properties of the accessibility relation

The modal logic K

If we define the logical properties this way and make no further restrictions on what counts as a model, we get the modal logic K. K is the weakest of the modal logics, and everything that is valid in K is valid in all the others. Remember the axioms, which can be proven to be tautologies based on the given semantics.

A1. $X \rightarrow (Y \rightarrow X)$
A2. $(X \rightarrow (Y \rightarrow Z)) \rightarrow ((X \rightarrow Y) \rightarrow (X \rightarrow Z))$

$$\text{A3. } ((\sim X) \rightarrow (\sim Y)) \rightarrow (Y \rightarrow X)$$

The axiom to specify distribution of necessity

$$\text{A4. } \Box(X \rightarrow Y) \rightarrow (\Box X \rightarrow \Box Y)$$

Here are some formulas that are logically true in K:

$$\Box(X \wedge Y) \equiv (\Box X \wedge \Box Y)$$
$$(\Box X \wedge \Box Y) \equiv \Box(X \wedge Y)$$
$$\neg\Box X \equiv \Diamond \neg X$$
$$\Box\neg X \equiv \neg\Diamond X$$

Can you see why they are true in all models? Think about (1) and (2) this way: if $X \wedge Y$ is true in all accessible worlds, then it must be that *X* is true in all those worlds, and *Y* is true in all those worlds. The converse also holds: if *X* is true in all accessible worlds, and so is *Y*, then $X \wedge Y$ is true in all accessible worlds.

Do you see the resemblance between (3) and (4) and the quantifier-negation equivalences? What explains this resemblance?

Here is a formula that is *not* logically true in K:

$$\Box X \rightarrow X$$

Here is an invalidating model:

R(w0,w1), I(w0,X)=f, I(w1,X)=t

Here *X* is false in w0, even though it is true at every world accessible from w0.

Exercise: Find a K model in which $\Box X \rightarrow \Diamond X$ is false.

The modal logic D

If you add $\Box X \rightarrow \Diamond X$ as an axiom schema to K, you get a stronger logic D. (Stronger in the sense that it has more logical truths and more valid arguments.)

Since D is stronger than K, there must be K-models that are not D-models. (This makes it easier to find counterexamples in K.) In fact, D-models are K-models that meet an additional restriction: the accessibility relation must be *serial*.

A relation *R* on *W* is *serial* iff $(\forall w \in W: (\exists w' \in W: (w,w') \in R))$

What this means, intuitively, is that there are no "dead ends"—no worlds that can't "see" any worlds (including themselves).

With dead ends ruled out, $\Box X \rightarrow \Diamond X$ no longer has counterexamples.

Note that on the deontic interpretation of the modal operators, where $\Box$ means "it is obligatory that" and $\Diamond$ means "it is permissible that," $\Box X \rightarrow \Diamond X$ is essentially the principle "ought implies can." So D is a good logic for this interpretation. Note that in a deontic logic, we don't want $\Box X \rightarrow X$, since often what ought to be the case isn't the case.

The modal logic T

If you add $\Box X \rightarrow X$ as an axiom schema to K, you get a stronger logic T.

A T-model is a K-model whose accessibility relation is *reflexive*. A relation R on W is *reflexive* iff ($\forall w \in W$: $(w,w) \in R$). That is, every world can see itself.

Since every reflexive accessibility relation is serial, every T-model is a D-model. The converse does not hold: there are D-models that are not T-models. Hence, every logical truth of D is a logical truth of T, there are logical truths of T that are not logical truths of D.

The modal logic S4

If you add $\Box X \rightarrow \Box\Box X$ as an axiom schema to T, you get a stronger logic S4.

An S4-model is a K-model whose accessibility relation is *reflexive* and *transitive*. A relation R on W is *transitive* iff ($\forall w1,w2,w3\ w \in W$: $(w1,w2) \in R$ & $(w2, w3) \in R \Rightarrow (w1,w3) \in R$). Theorems of S4 include $\Diamond\Diamond X \rightarrow \Diamond X$ and $\Diamond\Box\Diamond X \rightarrow \Diamond X$

The modal logic B

If you add $X \rightarrow \Box\Diamond X$ as an axiom schema to T, you get a stronger logic B. (Note that neither B nor S4 is stronger than the other; there are logical truths of B that are not logical truths of S4, and vice versa.)

A B-model is a K-model whose accessibility relation is *reflexive* and *symmetric*. A relation R on W is *symmetric* iff ($\forall w1,w2 \in W$: $(w1,w2) \in R \Rightarrow (w2,w1) \in R$)

The modal logic S5

If you add $\Diamond X \rightarrow \Box\Diamond X$ as an axiom schema to T, you get a stronger logic S5. S5 is stronger than both S4 and B.

An S5-model is a K-model whose accessibility relation is *reflexive*, *symmetric*, and *transitive*. That is, it is an *equivalence relation*.

It is easy to see that if a formula can be falsified by an S5-model, it can be falsified by a *universal* S5-model - one in which every world is accessible from every other. So we get the same logic if we think of our models as just sets of possible worlds. Because the accessibility relation is an equivalence relation, we can more or less forget about it, and talk of necessity as truth in all possible worlds. S5 is the most common modal logic used by philosophers, and semantically it is simpler than the others.

Exercise: Find an S5-model in which $\Diamond X \rightarrow \Box X$ is false.

The relationships between these six modal systems are summarized in the table bellow:

Property of relation	How the property is defined	Equivalent Axiom	System
Serial	iff ($\forall w \in W$: ($\exists w' \in W$: $(w,w') \in R$))	$\Box X \rightarrow \Diamond X$	**(D)**
Reflexive	iff ($\forall w \in W$: $(w,w) \in R$)	$\Box X \rightarrow X$	**(T)**
Transitive	iff ($\forall w1,w2,w3 \in W$: $(w1,w2) \in R$ & $(w2, w3) \in R \Rightarrow (w1,w3) \in R$)	$\Box X \rightarrow \Box\Box X$	**(4)**
Symmetric	iff ($\forall w1,w2 \in W$: $(w1,w2) \in R \Rightarrow (w2,w1) \in R$)	$X \rightarrow \Box\Diamond X$	**(B)**

Euclidian	iff	$(\forall w1,w2,w3 \in W: (w1,w2) \in R$ & $(w1, w3) \in R \Rightarrow (w2,w3) \in R)$	$\Diamond X \rightarrow \Box \Diamond X$	(5)

Theorem 5.1. The formula T : $\Box X \rightarrow X$ is a tautology (axiom) **iff** R is reflexive.

Proof.

1. If the axiom is a tautology then it is true in any world of a model. Suppose that R is not reflexive, then it exists a world w which is not accessible from itself. Suppose also that proposition X is true in all worlds except w. Then we have $\Box X$ is **true** in w, because the only world in which X is **false**, the world w, is not accessible starting from w. Therefore contradiction ($\Box X \rightarrow X$ is **false** in w).

2. If R is reflexive then w is accessible starting from itself. In these conditions, if $\Box X$ is **true** in w then X is also **true** in w, and the axiom $\Box X \rightarrow X$ has the value **true** in any world if R is reflexive.

Theorem 3.2. The formula $\Box X \rightarrow \Box\Box X$ is a tautology (axiom) **iff** R is transitive.

Proof.

1. Assume that $\Box X \rightarrow \Box \Box X$ is a tautology but R is not transitive. Then there are at least 3 worlds w, w' et w" such that wRw', w'Rw" and (w,w") $\notin$ R. Assume also that proposition X is **true** in any world of W except world w". Then $\Box X$ **is true** in w and **false** in w' (the world w" is accessible from w', but it is not accessible from w). Then the formula $\Box \Box X$ has the value **false** in w because $\Box X$ is **false** in w', therefore $\Box X \rightarrow \Box \Box X$ is **false** in w. Contradiction.
2. Assume that R is transitive. Let w be a world where formula $\Box X$ is **true** and w' a world accessible from w. Then X **is true** in w'. The formula X will be also true in all worlds w" accessible from w', because these worlds are accessible by transitivity from w (where $\Box X$ is **true**). Therefore $\Box X$ **is true** in any world w' accessible from w, which means that the formula $\Box \Box X$ has the value **true** in w, thus $\Box X \rightarrow \Box \Box X$ is **true** in any world if the relation R is transitive.

Theorem 3.3. The formula (5): $\Diamond X \rightarrow \Box \Diamond X$, is a tautology (axiom) **iff** R is Euclidian.

Proof.

1. Suppose that the formula $\Diamond X \rightarrow \Box \Diamond X$ is a tautology, then it is true in any world of the model, but the relation R is not Euclidian. Then there are at least 3 worlds w, w' and w" such that wRw', wRw" and (w',w") $\notin$ R. Assume also that the proposition X is **true** only in the world w". From wRw' and wRw" we can deduce that the formula $\Diamond X$ is **true** in w, but it is **false** in w' because (w',w") $\notin$ R. But w' is accessible from w (we have wRw'), then the formula $\Box \Diamond X$ is **false** in w (because $\Diamond X$ is **false** in w'). Therefore $\Diamond X \rightarrow \Box \Diamond X$ is false in w. Contradiction.

2. Assume that the relation R is Euclidian. If $\Diamond X$ is **true** in a world w then there exists a world w' accessible from w where X is **true**. But all worlds w" accessible from w, because of the Euclidian property, have access to world w'. This means that $\Diamond X$ is **true** in al these worlds w" which makes the formula $\Box \Diamond X$ **true**, therefore $\Diamond X \rightarrow \Box \Diamond X$ is **true** in all worlds of any model where the relation R is Euclidian.