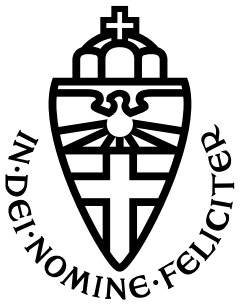
THE LIMITS OF APPLYING TEMPORAL LOGIC TO THE PROBLEM OF TIME

S. SCHAKENBOS

Supervisor N.P. Landsman

Abstract. Temporal logic is one form of modal logic which has shown its usefulness in many applications. Curiously, the problem of time in philosophy and physics is an issue where the literature is lacking in applying temporal logic. In this thesis temporal logic is extensively explained, and then applied to two interpretations of the problem of time: that of the philosopher and of the physicist. In philosophy we look at McTaggart's famous argument for the unreality of time. Using a new lemma which shows the effect of formulas which define the empty class to the consistency of a logic, the incompatibility of McTaggart's worldview and temporal logic is shown. In physics, a mathematical approach to (quantum) gravity called causal set theory is examined. The standard approach of a specialised logic is then shown to be incompatible with the specifics of causal set theory.



Radboud University
Date: September 2020

Contents

1. Introduction

> "What then, is time? If no one asks me, I know; but if I wish to explain it to someone who should ask me, I do not know."
>
> - St. Augustine, *Confessions*

This quote of St. Augustine captures the difficulty of time. While we are living our daily lives, we do not need to stop and think while using it to function in our highly time-regulated world. From the moment we wake up to when we go to bed, we are aware of the passage of time. We can use it to meet our appointments and spend it on some leisure. But if one is asked to explain what time is, almost no one would have an answer. Most likely one would say: "Time is... of course... you know. Time," after which they would fall silent.

If one were to ask a more philosophically inclined person, they might come with something a bit more coherent. "Time moves forward, which we can see from the change all around us." And while an understandable answer, even satisfying for some people, it only tells us what we observe of time, not what it is. But maybe we can ask some experts.

First we might turn to philosophy, and ask some individual philosopher who has studied such metaphysical concepts. But while he can probably talk at lengths how he sees the universe, for a consensus he must admit that, like every question in philosophy, there is only a fierce debate.

Then we might turn to the people who deal with time in their daily work, physicists. But we will encounter a familiar scene. While those whose work mainly lays in quantum theory will explain to you that time is a background parameter along which change can occur, those who work with general relativity warn that time is not something separate, but is intricately linked with space and its contents within the larger spacetime structure.

And if we ask those who are working at the frontier of science, the study of quantum gravity, they will laugh, and say that if you can answer that question you will probably receive a Nobel price.

This was the state of our understanding of time as I stepped into this thesis. There is almost nothing we actually know about time, and there is much discussion about even how much can be known about time. Is the past even real? What about the future? Can we even trust our mind about the present. Biology tells us that our feeling of the presence is made up from what has happened in the last 80 milliseconds, which gives doubts about our physical observations.

These questions about the nature of time are generally put under the philosophical problem of time, which is a field within meta-physics. This is in contrast to the physical problem of time, which specifically refers to the issue of time within the giant problem that is quantum gravity. It is a coincidence that they use the same name, as internally they both just call it the problem of time. But in that light the issues arose to which we are going to apply temporal logic to.

Why temporal logic? It appears that temporal logic, being an non-standard logic, has not yet been applied to these problems. This might first sound odd, as temporal logic seems made to describe these issues. But when Arthur Prior introduced temporal logic in the early 1950s as a version of the more general modal logic first introduced by C.I. Lewis in 1918, it was to describe the interaction of propositions with temporal concepts. The founder Prior, among other mathematicians, studied temporal logic as in its own right. And while the syntax of temporal logic was adopted

by philosophers, its logical properties were almost never used to prove aspects about reality.

This explains a bit why temporal logic had not yet been applied to the philosophical problem of time, but when we look at physics, it is more likely a lack of familiarity on both sides. Very few mathematicians know enough about the physical problem of time to see any connection. And for most physicists there are more enticing mathematical theories to first apply.

So while the application being unknown territory seemed promising at the start, it was unfortunately that my research had resulted in negative proofs. By studying the problem of time in both philosophy and physics I encountered two instances where the application of temporal logic seemed fitting, but which in the end came back with showing in what ways temporal logic is lacking.

Before we discuss the results, it is required to explain temporal logic. As a bachelor student, one usually is only introduced to propositional logic and first order logic, where the former is usually only a stepping stone to get to the latter. Personally I feel that this does not do justice to the diversity of formal logic, so before I look at modal logic, I will describe formal logic not through an example, but in general. Then the theory of propositional logic is described in great detail, as it will help to directly see the contrast between propositional logic and modal logic.

Temporal logic, while very interesting on its own, is only one of the many schools of logic that all fall under modal logic. The theory of modal logic is not something one encounters in a bachelor programme, so it will be explained in great detail. Then the specific application of modal logic in the form of temporal logic will be surprisingly short, as most of the work has already been done in the previous section.

The two new results of this thesis are built upon a new lemma. When we are looking at a logic we generally ask if it is sound and complete. But in modal logic it is more common to ask to what extent a logic is sound and complete. We show in lemma 4.27 that a logic which is extended by a certain type of formula will become inconsistent, which is the lowest level of soundness and completeness.

Turning to the philosophical problem of time, we will focus on one of the most important arguments of the entire philosophy of time. Introduced by McTaggart, it spawned two opposing camps on the way we should fundamentally think about time. In theorem 6.2 the lemma 4.27 us then used to show that the world according to McTaggart cannot be governed by the rules of Temporal Logic.

The second result, that of the physical problem of time, has a less definitive result, but still a limit on the applicability of temporal logic. One of the approaches to the study of (quantum) gravity, causal set theory, describes reality very similar to the way temporal logic describes reality. But in theorem 7.7 we show that the typical approach to generating a logic which is sound and complete to the extent wanted is impossible.

Thus in section 2, formal logic is described. Then in section 3 the theory of propositional logic is revisitet, and in section 4 we find an overview of the important properties of modal logic. This theory is then extended to the more specific temporal logic in section 5. Section 6 then applies temporal logic to the philosophical problem of time, and section 7 applies it to the physical problem of time.

2. What is Formal Logic?

For most students of mathematics their first lecture will be about the fundamentals of mathematics: formal logic. Most of the time, formal logic is explained through the example of propositional logic, or sometimes first order logic. And there are good reasons for it. Didactically it is better to slowly introduce the new concept, and couple it back to knowledge the students already have. The ideas of formal logic are not simple, as it covers many different ways to think about mathematical reasoning. But an unfortunate side-effect of this focus on propositional logic is that the true scope of formal logic remains hidden to most students.

So the question remains, what is formal logic. Succinctly put, a logic is a set of rules that describe the *syntax* with one algorithm to determine truth, *semantics*, and one algorithm to determine theoremhood, *deduction*.

The syntax of a logic prescribes what a valid expression is. While an expression is a finite sequence of arbitrary symbols, a (well-defined) formula is an expression that adheres to the syntax. While this is enough to define what syntax is, most of the time it has a more definite form. It is usually defined through some form of iteration, with a finite set of rules. Otherwise the scope of the language becomes so big that it is difficult to say anything meaningful. Furthermore, it is common for a logic to speak about the relations between an indefinite number of objects. To represent this, there is usually a countable set of symbols which have the same minimal role, except for being distinct from each other, called variables. These variables usually start free, but through certain syntax rules can become bound.

The semantics is an algorithm which maps formulas to a set called the truth values. Again, this is the minimal definition, but usually the algorithm is closely linked to the iteration of the syntax. If a syntax rule states that two smaller formulas can be combined, then the semantic algorithm will tell you how the truth value of the larger formula is derived from the truth values of the smaller components. But most of the time, the algorithm cannot tell us on its own what the truth value of a formula is, or else there cannot be any conditionality truth. That smallest structure that fully determines the semantics is called a model. And in semantics with models, those formulas that have a model-independent truth value are of special interest, since they represent something absolute in the changing truth of the semantics.

Deduction is an algorithm to determine the theoremhood, whether something is a theorem, of sentences, formulas with only bound variables. It is defined in the terms of syntax, independently of the semantics.

The second kind of algorithm, deduction, is less fundamental, as it is made to fit nicely with the syntax and the semantics. It is defined in the terms of the syntax, independently of the semantics. The algorithm, while usually not completely understood, maps formulas to the two values of provable or not provable. Unlike with the syntax and semantics, there is not much more unity in how deduction is formatted, except in how it interacts with semantics.

The way deduction and semantics are linked is usually expressed in soundness and completeness theorems. A sound logic means that there is a deduction system where all provable formulas are true model-independently. A complete logic means that there is a deduction system which proves all formulas that are model-independently true. For any logic it is a goal to have the semantics and deduction be so that the logic is sound and complete.

In practice, a new logic will be created to formalise the interaction between a set of expressions with an underlying structure. A syntax is set up in order to rigorously define what expressions are possible. Then either the intuitive understanding of the truth values of different expressions is formalised in a semantic algorithm, or the intuitive interaction is formalised in a deduction. By searching for an appropriate counterpart such that the logic is at least sound, and as close to complete as possible, it is hopefully possible to express all the properties of the underlying structure.

2.1. **An Unusual Example.** As already mentioned above, the usual (first) example that is used to illustrate the fundamentals of formal logic is propositional logic. But a better example for the process described at the end of the section might be to use a new logical language. So let me introduce a language many of the readers will already be familiar with: algebraic chess notation.

Algebraic chess notation is the name for how the official world chess federation FIDE prescribes that a chess game should be recorded. While going into the details is not the interesting part, a short description in necessary[1]. A game of chess is defined by the sequence of moves played, so the syntax of chess consists of two columns of numbered rows, each consisting of a single move. A move indicates which piece moves where, and whether it captures a piece or not. There are some additional decorations that make it easier to read, but these are not part of the underlying logic of valid chess games.

The syntax of algebraic chess notation is a long, but finite list of ways a game can be started or extended. Since (almost) every piece can move to every square, a well defined chess game is any numbered list of moves which does not include any impossible moves. A chess game is then generated by iterating these extensions, until a completed game is reached. Chess does not have a set of infinite symbols, which is fine. Algebraic chess notation is not intended to describe something with an infinite number of objects in it, so it is fine that there is no set of self-similar but distinct symbols.

The semantics of algebraic chess notation should encompass our understanding of what makes a game real, i.e. whether it is actually possible to play it. For example, if the notation tells us to move a piece to a square it can't reach, that would tell us that the game is not possible. This is captured with two truth values: a game is either possible or it is not possible. Now it is important to see that, whether a move like Nf3 (move the knight to the square f3) is a possible move, requires us to know what the piece's original position was. Most of the time the position of a piece can be extracted from the game with the last move removed. We see that the semantics indeed follows the iteration of the syntax. But there is one situation where this extraction fails, namely if the piece is still in its starting position. So the starting positions of all the pieces are the minimum required information necessary to determine the truth value of any chess game. The starting positions are the models of algebraic chess notation. The most common model would be the default starting position, but different starting positions are also possible (for example 960chess[2]).

This leaves us with the challenge of finding a deduction system which is sound and complete to algebraic chess notation. But unfortunately, this thesis was not about algebraic chess notation, and I can therefore not give the answer. But there are still

[1]For the notation see `https://www.fide.com/FIDE/handbook/LawsOfChess.pdf` Appendix C.

[2]For an official explanation, see `https://www.fide.com/FIDE/handbook/LawsOfChess.pdf` Appendix F.

interesting aspects algebraic chess notation has that can be easily understood. For example, there is no game that is model-independently possible, so we are doomed if we want to find a sound and complete deduction system. But by being a bit less strict we can get some interesting results. If we have a deduction system that proves "1. e4" (first move of the game, move a white pawn to the square e4), then it is not sound, since our logic has models such that the game "1. e4" is invalid. But there are models for which "1. e4" is a valid game, and maybe we can say something meaningful about those models as a sub class of all the models in algebraic chess notation.

Hopefully this section has given a deeper understanding of what formal logic is. While it will not be referred to in the rest of the thesis, it is still important to the fundamental understanding I am trying to convey.

3. Propositional Logic

Propositional logic is the basis for many other logic systems, which is why we will spend a lot of space discussing it. We will first introduce the syntax, after which we will treat both semantic truth evaluation and deduction. Then in section 3.4 we show the interaction between the semantics and the deduction through the fundamental theorem of propositional logic. We end with some closing remarks. Most of this section is adapted from N.P. Landsman, *Propositielogica* (in Dutch) [1].

3.1. Syntax of Propositional Logic.

The first components in propositional logic are encoded in the **signature** $S = \{p_1, p_2, \dots\}$[3]. This is an infinite set, but the letters p, q, r, s are also commonly used. These symbols represent the **atomic propositions**, which are the smallest expressions that can form a formula, as explained below. Atomic propositions can be stitched together to form longer formulas using connectives. These connectives are *or* ($\vee$), *and* ($\wedge$), *right implication* ($\to$) and *bi-directional implication* ($\leftrightarrow$). There are additional symbols for *not* ($\neg$), *falsum* ($\bot$) and *truth* ($\top$). Lastly, we have the brackets, which are used to distinguish between ambiguous formulas. Most of the time, they can if $\neg$ binds stronger, then $\vee$ and $\wedge$, followed by $\to$ and $\leftrightarrow$. The symbols $\top$ and $\bot$ take the same place as atomic propositions, so they remain outside this hierarchy.

We first restrict ourselves to a smaller set of symbols: the elements of S, $\bot$ and $\to$[4]. This is done to streamline later proofs, as the other symbols will be defined as abbreviations of formulas. With this smaller set we define **(well-defined) formulas** as follows.

Definition 3.1. A series of symbols is a well-defined formula if it can be constructed by iterating the following three rules a finite number of times:

(1) Each symbol $p \in S$ is a well-defined formula.
(2) $\bot$ is a well-defined formula.
(3) If φ and ψ are well-defined formulas, then $\varphi \to \psi$ is a well-defined formula.

The set of all well-defined formulas is denoted by $\mathcal{L}_S$.

[3]It is possible to work with an uncountable signature, but since in this thesis we will strictly be dealing with finite formulas, there is no value in having an uncountable number of atomic propositions.

[4]There are many choices as to which symbols we take to be the base ones, and the ones chosen here is merely a matter of preference. Technically, the *not and* (NAND) connective would be sufficient on its own, but this is not chosen for clarity.

Symbol	Abbreviation
$\neg p$	$p \to \bot$
$p \vee q$	$\neg p \to q$
$p \wedge q$	$\neg(\neg p \vee \neg q)$
$p \leftrightarrow q$	$(p \to q) \wedge (q \to p)$
$\top$	$\bot \to \bot$

TABLE 1. Symbols and their abbreviations in propositional logic.

So the formula $p \to \bot$ is well-defined formula, since it can be constructed by first using rule (1) and (2) to get the formulas p and $\bot$. Then using rule (3) we combine the two formulas into one larger formula $p \to \bot$.

The other symbols are used as abbreviations, as shown in table 1.

3.2. **Semantics of Propositional Logic.** When we look at semantics and deduction of a logic, one needs to be chosen as the more fundamental one. For the other a system is sought which plays nice with the fundamental one. Which is chosen as fundamental is a matter of preference.

In mathematics it is common to describe the semantics before the deduction of a logic. Of the two, the semantics is the more fundamental one. Usually a deduction system is sought to play nicely with the semantics, and not the other way around.

3.2.1. *Models and Evaluations.* When reading the formulas through a semantic lens, the atomic propositions are the smallest formulas with their own meaning of truth. They can be either true or false. These two **truth values** are represented by the numbers 1 and 0, respectively. Examples of such smallest formulas are *"Socrates had a beard"*, *"The grass is green"* and *"It is raining now"*. What the truth value is of these expressions will depend on what the facts are, which is captured in a **model**.

Definition 3.2. A model consists of what truth value each atomic proposition has. It is encoded in its evaluation function $val : S \to \{0, 1\}$.

It is possible to extend a model into an evaluation function on the whole $\mathcal{L}_S$.

Lemma 3.3. *Each evaluation val can be extended to a function* $Val : \mathcal{L}_S \to \{0, 1\}$ *with the following definition:*

$$\begin{aligned} Val(p) &= val(p) \quad \textit{for each } p \in S \\ Val(\bot) &= 0 \\ Val(\varphi \to \psi) &= Val(\varphi) \to' Val(\psi) \end{aligned}$$

Where the symbol $\to'$ *refers to the binary operator shown in table 2.*

Existence and uniqueness follows from the correspondence between the way the evaluation is extended an the rules for constructing well-defined formulas. Similar to the way the $\to$ is treated, the other connectives are pulled out:

$$Val(\varphi \vee \psi) = Val(\varphi) \vee' Val(\psi).$$

It is left as an exercise to the reader to show that every abbreviation mentioned in table 1 match up with their counterparts in table 2, and that $Val(\top) = 1$.

An additional way to look at $Val(\varphi)$ is as an function, which maps the different values $val(p_i)$ can take to the truth values $\{0, 1\}$. Such function can implicitly be defined:

a b	a $\rightarrow$' b	$\neg$' a	a $\vee$' b	a $\wedge$' b	a $\leftrightarrow$' b
0 0	1	1	0	0	1
0 1	0	1	1	0	0
1 0	1	0	1	0	0
1 1	1	0	1	1	1

TABLE 2. Behaviour of binary counterparts to $\rightarrow, \neg, \vee, \wedge$ and $\leftrightarrow$.

the function form of φ is a function φ' such that $Val(\varphi) = \varphi'(val(p_1), \ldots, val(p_n))$. The mapping $\varphi \mapsto \varphi'$ is unique, since the extension from val to Val is unique.

3.2.2. *Tautologies.* While looking at individual models can lead to interesting results, more fundamental results are obtained if we look at the formulas whose truth value is independent of what model, i.e. which function val we are using. It is always possible to check whether a formula φ has a universal truth value, since we only need to check a finite number of models. Since any formula consists of a finite number n of atomic propositions, there can at most be 2^n different val that need to be checked.

Definition 3.4. When a formula φ has a universal truth value, and that value is 1 (or true), we call φ a **tautology**. We write this as $\models \varphi$.

Let us look at the formula $\varphi = p \rightarrow p$. Since we have one atomic proposition, we already know that we only need to check two evaluations to determine if φ is a tautology. So we have either $val(p) = 1$ or $val(p) = 0$. Now since $Val(\varphi) = val(p) \rightarrow' val(p)$, and both the 0 0 and the 1 1 row of table 2 have a $\rightarrow$' b be equal to 1, we can conclude $\models \varphi$.

Now let us show a simple result.

Lemma 3.5. *If $\models \varphi$ and φ has n atomic propositions and we have a set $\Sigma = \{\alpha_1, \ldots \alpha_n\}$ of formulas in $\mathcal{L}_S$, then $\models \varphi^*$, where φ^* is obtained by replacing the instance of p_i in φ with α_i.*

An example, we have already shown $\models p \rightarrow p$, so as a corollary of this lemma, we can conclude that, for each $\beta \in \mathcal{L}_S$, $\models \beta \rightarrow \beta$.

Proof. Assume that φ^* is not a tautology. Then there is a model val^* such that $Val^*(\varphi^*) = 0$. Let val be such that $val(p_i) = Val^*(\alpha_i)$. Then

$$Val(\varphi) = \varphi'(val(p_1), \ldots, val(p_n)) = \varphi'(Val^*(\alpha_1), \ldots, Val^*(\alpha_n)) = Val^*(\varphi^*).$$

So $Val(\varphi) = 0$, which means that φ is not a tautology, which is a contradiction. Thus $\models \varphi^*$. □

3.2.3. *Semantic Deduction.* It is clear that tautology is a strong attribute. Sometimes, it is useful to limit the evaluations we examine. Let us extend the notation $\models$ to encompass this idea. Let $\Sigma = \{\alpha_1, \ldots, \alpha_n\}$ be a set of formulas in $\mathcal{L}_S$. Such a set is called a **theory** We say that $Val(\Sigma) = 1$ if and only if $Val(\alpha_i) = 1$ for each $\alpha_i \in \Sigma$. Then Σ semantically implies φ, noted as $\Sigma \models \varphi$, if and only if for each evaluation val such that $Val(\Sigma) = 1$, we have that $Val(\varphi) = 1$. The case $\{\alpha\} \models \varphi$ will be abbreviated to $\alpha \models \varphi$. In the case that there are no evaluations such that $Val(\Sigma) = 1$, e.g. both p and $\neg p$ are elements of Σ, we find that $\Sigma \models \varphi$ for each $\varphi \in \mathcal{L}_S$. Especially $\Sigma \models \perp$ if and only if there are no evaluations such that $Val(\Sigma) = 1$.

We can now formulate the, semantic deduction theorem, which is similar to the deduction theorem (theorem 3.11) but in a semantic context.

Theorem 3.6. $\Sigma \models \alpha \to \beta$ *if and only if* $\Sigma \cup \{\alpha\} \models \beta$.

Proof. The proof is a direct consequence of the a →' b column of table 2. We have a →' b be true if and only if a = 0 or b = 1. So we have $Val(\alpha \to \beta) = 1$ if and only if $Val(\alpha) = 0$ or $Val(\beta) = 1$. To show the **if**, we assume $\Sigma \models \alpha \to \beta$. So for each *val* where $Val(\Sigma) = 1$, we have $Val(\alpha) = 0$ or $Val(\beta) = 1$. But when we are looking at $\Sigma \cup \{\alpha\} \models \beta$, we need to disregard those *val* which have $Val(\alpha) = 0$. So we only look at those *val* such that $Val(\beta) = 1$, but then necessarily we have $Val(\alpha \to \beta) = 1$, and thus $\Sigma \cup \{\alpha\} \models \beta$.

To show the **only if**, we assume $\Sigma \cup \{\alpha\} \models \beta$. Since $\Sigma \cup \{\alpha\} \models \beta$ implies $\Sigma \cup \{\alpha\} \models \alpha \to \beta$, because $Val(\beta) = 1$, we only need to consider those *val* such that $Val(\Sigma) = 1$ and $Val(\alpha) = 0$. But then necessarily we have $Val(\alpha \to \beta) = 1$. And thus $\Sigma \models \alpha \to \beta$. □

3.3. **Deduction in Propositional Logic.** We will now focus on deduction, and while this is closely related to semantics, this section will focus on what deduction is in its own right. In the next section we will see how these two notions interact with each other. While we will be discussing deduction within propositional logic, most of the ideas introduced here can be generally applied.

Deduction is a method that allows us to determine the theoremhood of formulas in propositional logic. When a formula φ is deducible, then we denote it as $\vdash \varphi$, and φ is called a **theorem**.

3.3.1. *The Deduction System of Propositional Logic.* A deduction system consists of two components, a (not necessarily finite) set of formulas called axioms, and a list of rules which we can apply, which are called deduction rules. While in theory any deduction system could be used, as we will see in section 3.4 there will be some strict requirements put on what deduction systems we deem acceptable. To make deduction more rigorous, let us use one possible definition of deduction systems and proofs.

Definition 3.7. A deduction system, or just system, is a pair (A, D), where A is a set of formulas in $\mathcal{L}_S$ called the axioms, and $D = \{d_1, d_2, \dots\}$ is a countable set of finite algorithms which take n_i formulas of a certain form, called the input, and returns a new formula called the output.

Definition 3.8 (Proofs in Propositional Logic)**.** A proof of φ within a system (A, D), given the theory Σ as assumptions, is a finite numbered list of formulas $(\alpha_1, \dots, \alpha_N)$ such that $\alpha_N = \varphi$ and for each α_i, either $\alpha_i \in A \cup \Sigma$ or there is a rule $d \in D$ and a subset of $\{\alpha_1, \dots, \alpha_{i-1}\}$ which fits the requirements of the input of d such that α_i is the output.

We then say that $\Sigma \vdash \varphi$ (in the system (A, D)) if there exists a proof of φ with the assumptions Σ. We then abbreviate $\{\alpha\} \vdash \varphi$ and $\emptyset \vdash \varphi$ as $\alpha \vdash \varphi$ and $\vdash \varphi$ respectively.

Even with the choice of what a deduction system is, there are many equivalent deduction systems for propositional logic. For this thesis, we will use A. Church's axioms, but this is not the only option.

Definition 3.9 (A. Church's deduction system)**.** There is one deduction rule and an infinite number of axioms. These axioms can take one of three forms, defined via axiom schemes. From the schemes, an axiom is obtained by replacing α, β and γ with any three formulas from $\mathcal{L}_S$

Deduction rule:

Modus Ponens: As input we take two formulas of the forms α and $\alpha \to \beta$ and we output β.

Axiom schemes:

(1) $\beta \to (\alpha \to \beta)$
(2) $\big(\beta \to (\alpha \to \gamma)\big) \to \big((\beta \to \alpha) \to (\beta \to \gamma)\big)$
(3) $\big((\alpha \to \bot) \to (\beta \to \bot)\big) \to \Big(\big((\alpha \to \bot) \to \beta\big) \to \alpha\Big)$

Axiom scheme 1 and 2 regulate the $\to$ symbol, while axiom scheme 3 regulates the $\bot$ symbol. Note that all axioms obtained from these schemes are tautologies. Why this is the case is commented upon in section 3.5.

Let us look at an example of how a proof would look.

Example 3.10. One proof for $\vdash \alpha \to \alpha$ is

(1) $\alpha \to ((\alpha \to \alpha) \to \alpha)$.
(2) $(\alpha \to ((\alpha \to \alpha) \to \alpha)) \to ((\alpha \to (\alpha \to \alpha)) \to (\alpha \to \alpha))$.
(3) $(\alpha \to (\alpha \to \alpha)) \to (\alpha \to \alpha)$.
(4) $\alpha \to (\alpha \to \alpha)$.
(5) $\alpha \to \alpha$.

This is a valid proof because of the explanation:

(1) Use Axiom 1 where β is replaced with α and α is replaced with $(\alpha \to \alpha)$
(2) Use Axiom 2 where both β and γ is replaced with α and α is replaced with $(\alpha \to \alpha)$.
(3) Use the Modus Ponens rule with input formulas 1 and 2.
(4) Use Axiom 1 where both α and β are replaced with α.
(5) Use the Modus Ponens rule with input formulas 3 and 4.

3.3.2. *Deduction Theorem.* There is an important relation between the Modus Ponens deduction rule and the theory one uses in a proof. This is captured in the deduction theorem:

Theorem 3.11 (Deduction Theorem)**.** $\Sigma \vdash \alpha \to \beta$ *if and only if* $\Sigma \cup \{\alpha\} \vdash \beta$.

Proof. The **if**: since $\Sigma \vdash \alpha \to \beta$, we also have $\Sigma \cup \{\alpha\} \vdash \alpha \to \beta$. By adding to the proof the assumption α and using Modus Ponens, we get the output β. Thus $\Sigma \cup \{\alpha\} \vdash \beta$.

The **only if**: We use induction on the length of the proof of $\Sigma \cup \{\alpha\} \vdash \beta$. If it is of length one, β has to be an element of $\Sigma \cup A$. Use the Modus Ponens rule on β and Axiom 1, $\beta \to (\alpha \to \beta)$ to get $\alpha \to \beta$.

Assume the deduction theorem is true for proofs of length n or less, and that $\Sigma \cup \{\alpha\} \vdash \beta$ is a proof of length $n+1$. Were $\beta \in \Sigma \cup A$, we can use the same argument as above. If it is not the case, then β is concluded from the Modus Ponens rule. Therefore there are two shorter sub-proofs such that $\Sigma \cup \{\alpha\} \vdash \gamma$ and $\Sigma \cup \{\alpha\} \vdash \gamma \to \beta$ for some $\gamma \in \mathcal{L}_S$. Now we apply the inductions hypothesis to conclude $\Sigma \vdash \alpha \to \gamma$ and $\Sigma \vdash \alpha \to (\gamma \to \beta)$. Using axiom 2,

$$(\alpha \to (\gamma \to \beta)) \to ((\alpha \to \gamma) \to (\alpha \ \to \beta)),$$

and the Modus Ponens rule twice, we find $\Sigma \vdash \alpha \to \beta$.

Therefore we can conclude $\Sigma \vdash \alpha \to \beta$ if and only if $\Sigma \cup \{\alpha\} \vdash \beta$. □

The deduction theorem is a powerful tool for quickly and succinctly proving many useful results. For example, it is possible to show that from $\vdash \alpha \to (\beta \to \gamma)$ and $\vdash \beta$ we can conclude $\vdash \alpha \to \gamma$ without the need of any of the axioms. The proof itself is left as an exercise to the reader.

3.3.3. *Consistency.* While any subset of $\mathcal{L}_S$ is a valid theory, many will not be useful, because they can prove any formula. We call a theory Σ **inconsistent** if $\Sigma \vdash \bot$. Because $\bot \vdash \varphi$ for every formula φ, we find that if Σ is inconsistent, $\Sigma \vdash \varphi$. Showing that $\bot \vdash \varphi$ is left as an exercise to the reader. Although for any theory that is intended to be used as a basis of a field of study, this would be detrimental, using inconsistent theories for proofs as those in section 3.4 is a useful tool. The counterpart of an inconsistent theory is a **consistent** theory, which is denoted by $\Sigma \nvdash \bot$. Now let us finish this section with a simple result that will be used in the next section.

Lemma 3.12. *Let $\Sigma \subset \mathcal{L}_S$ be a consistent theory, and let $\alpha, \beta \in \mathcal{L}_S$ be any two formulas.*

(1) *If $\Sigma \vdash \alpha$ then $\Sigma \cup \{\alpha\}$ is consistent.*
(2) *If $\Sigma \cup \{\beta\}$ is inconsistent, we have $\Sigma \vdash \neg\beta$, which with (1) means that $\Sigma \cup \{\neg\beta\}$ is consistent.*

Proof. For 1, adding a provable formula to a theory will not allow us to prove new formulas, because any proof that requires α as an assumption can be gotten by replacing that assumption with the proof $\Sigma \vdash \alpha$. Since the formula $\bot$ can't be proven from Σ, it can't be proven from $\Sigma \cup \{\alpha\}$, which is thus consistent.

For 2, if $\Sigma \cup \{\beta\}$ is inconsistent, there is a proof such that $\Sigma \cup \{\beta\} \vdash \bot$. Then, using the deduction theorem 3.11, we have $\Sigma \vdash \beta \to \bot$, which is $\Sigma \vdash \neg\beta$. □

3.4. **Soundness and Completeness.** While the semantics and the deduction of a logic are defined separately, they are generally chosen so that they have a nice interaction. The two big properties that a logic can have is that is being sound and complete.

Definition 3.13. A logic is **sound**, if every formula that is provable is semantically true, and a logic is **complete** if every semantically true formula is provable.

That propositional logic is sound and complete is a corollary of the fundamental theorem of propositional logic.

Theorem 3.14 (The Fundamental Theorem of Propositional Logic)**.** *For every signature S and every theory $\Sigma \subset \mathcal{L}_S$ we have that $\Sigma \vdash \varphi$ if and only if $\Sigma \models \varphi$. In particular, a formula is a tautology if and only if it is provable: $\vdash \varphi$ if and only if $\models \varphi$.*

3.4.1. *Why Soundness and Completeness?* The majority of this section is devoted to proving theorem 3.14, but let us first discuss its significance. What would a logic look like that isn't sound? Then there were some formulas which we deem to be false, but are provable. That would erode the intuitive meaning of being true. While for most mathematicians it is possible to distance themselves from the ordinary definition of words, and to look only at their semantic meaning, but whenever a logic is required to function when applied to reality, it seems necessary for it to be sound.

While soundness can almost be seen as a necessity for any "real" logic, completeness is at most an extremely nice feature. If it is shown that any true formula can be

proven, we know it is not in vein that we are looking for a proof. And as it turns out, both propositional logic and modal logic have this useful property.

3.4.2. *General Soundness.* The proof of theorem 3.14 will be done in two parts. First we will look at the comparatively easy soundness of propositional logic.

Lemma 3.15. *Propositional logic is generally sound, i.e. for every signature S and every theory $\Sigma \subset \mathcal{L}_S$, if $\Sigma \vdash \varphi$ then $\Sigma \models \varphi$.*

Proof. The proof is done by induction on the length of the proof of φ. Let $\Sigma \vdash \varphi$ be a proof of length one. That is only possible if $\varphi \in \Sigma \cup A$, where A are the axioms mentioned in definition 3.9. Now to show that $\Sigma \models \varphi$, we have to check all evaluations val such that $Val(\Sigma) = 1$. But by definitions, if $\varphi \in \Sigma$, we already have $Val(\varphi) = 1$. And since all axioms are tautologies, where $\varphi \in A$ it would also be the case that $Val(\varphi) = 1$. So $\Sigma \models \varphi$ must be true.

Now for a proof of length $n + 1$, the last formula has to be φ. This is either an element of $\Sigma \cup A$, which we have already proven, or there are two formulas α_i and α_j such that $i, j < n + 1$ and $\alpha_j = \alpha_i \to \varphi$ such that φ can be concluded because of the Modus Ponens deduction rule. Since it must be the case that $\Sigma \vdash \alpha_i$ and $\Sigma \vdash \alpha_i \to \varphi$ with proofs of length n or less, we can apply the induction hypothesis to get $\Sigma \models \alpha_i$ and $\Sigma \models \alpha_i \to \varphi$. So, examining the val such that $Val(\Sigma) = 1$, we have that $Val(\alpha_i) = 1$ and $Val(\alpha_i \to \varphi) = 1$. This last one requires, looking at the a→'b column of table 2, that either $Val(\alpha_i) = 0$ or $Val(\varphi) = 1$. And since $Val(\alpha_i) = 1$ the only case is $Val(\varphi) = 1$.

With induction we can conclude that if $\Sigma \vdash \varphi$ then $\Sigma \models \varphi$. □

With this, we have shown the easier of the two direction of the fundamental theorem of propositional logic. Now we will focus on the more difficult of the two.

3.4.3. *General Completeness.*

Lemma 3.16. *Propositional logic is generally complete, i.e. for every signature S and every theory $\Sigma \subset \mathcal{L}_S$ we have that if $\Sigma \models \varphi$ then $\Sigma \vdash \varphi$.*

First we will show that this lemma can be reduced to a simpler form.

Lemma 3.17. *If $\Sigma \models \bot$ then $\Sigma \vdash \bot$. By taking the contrapositive, this can be rephrased as: if Σ is consistent, then Σ has a model.*

Proof of reduction. Let us assume $\Sigma \models \varphi$. It is trivial to show that $\varphi \models \neg\varphi \to \bot$, so $\Sigma \models \varphi$ implies $\Sigma \models \neg\varphi \to \bot$. Using the semantic deduction theorem 3.6, we find $\Sigma \models \neg\varphi \to \bot$ implies $\Sigma \cup \{\neg\varphi\} \models \bot$. Now we used that $\Sigma \models \bot$ implies $\Sigma \vdash \bot$ to find $\Sigma \cup \{\neg\varphi\} \vdash \bot$. Using the deduction theorem 3.11, we get $\Sigma \vdash \neg\varphi \to \bot$. Then, using $\vdash (\neg\varphi \to \bot) \to \varphi$ and Modus Ponens rule, we can conclude $\Sigma \vdash \varphi$. Showing $\varphi \models \neg\varphi \to \bot$ and $\vdash (\neg\varphi \to \bot) \to \varphi$ is left as an exercises to the reader. □

To prove Lemma 3.17, we will first introduce the notion of a maximal consistent theory.

Definition 3.18. A maximal consistent theory is a consistent theory $\Sigma_m \subset \mathcal{L}_S$, such that if Σ is consistent and $\Sigma_m \subset \Sigma$, then $\Sigma_m = \Sigma$.

We will show two properties a maximal consistent theory has.

Lemma 3.19. *A maximal consistent theory Σ_m is deductively closed and complete, i.e. for each $\beta \in \mathcal{L}_S$ we have $\Sigma_m \vdash \beta$ implies $\beta \in \Sigma_m$ (deductively closed) and $\Sigma_m \vdash \beta$ or $\Sigma_m \vdash \neg\beta$ (complete).*

Proof. Deductively closed: from $\Sigma_m \vdash \beta$ we can conclude $\Sigma_m \cup \{\beta\}$ is consistent from Lemma 3.12.1. And because $\Sigma_m \subset \Sigma_m \cup \{\beta\}$, we have the maximal property, and thus $\beta \in \Sigma_m$.

Complete: if neither $\Sigma_m \vdash \beta$ nor $\Sigma_m \vdash \neg\beta$, then also neither β nor $\neg\beta$ are elements of Σ_m, or else there would be a trivial proof. Now, because of Lemma 3.12.2, it has to be the case that either $\Sigma_m \cup \{\beta\}$ or $\Sigma_m \cup \{\neg\beta\}$ is consistent. But neither can be the case, or else there would be a strictly larger consistent theory containing Σ_m. Therefore, we must have either $\Sigma_m \vdash \beta$ or $\Sigma_m \vdash \neg\beta$ □

Now we can go ahead and prove Lemma 3.17.

Proof of Lemma 3.17. We need to find, for a given consistent theory Σ, an evaluation *val* such that $Val(\Sigma) = 1$. We will directly define the function Val, and then show it is indeed an (extended) evaluation. This is done through the maximal consistent theory Σ_∞.

Because the signature is a countable set, it is also the case that $\mathcal{L}_S = \{\alpha_1, \alpha_2, \dots\}$ is a countable set[5]. Let us set $\Sigma_0 = \Sigma$. Then if $\Sigma_n \cup \{\alpha_{n+1}\}$ is consistent, set $\Sigma_{n+1} = \Sigma_n \cup \{\alpha_{n+1}\}$. If is not consistent, we define $\Sigma_{n+1} = \Sigma_n \cup \{\neg\alpha_{n+1}\}$. Now, each Σ_n is consistent because of Lemma 3.12.2. Then let $\Sigma_\infty = \bigcup_{n\in\mathbb{N}} \Sigma_n$. The claim is that Σ_∞ is a maximal consistent set.

First, let us show that Σ_∞ is consistent. Assume the opposite is true, so there is a proof $\Sigma_\infty \vdash \bot$. Since a proof is finite, only a finite number $\{\alpha_{n_1}, \alpha_{n_2}, \dots, \alpha_{n_k}\} \subset \Sigma_\infty$ are used in the proof. Let $N = \max_k\{n_k\}$, then $\Sigma_N \vdash \bot$. But that would mean Σ_N is inconsistent, which we know isn't the case. Therefore, Σ_∞ is consistent.

Now, assume there is a consistent Σ' such that $\Sigma_\infty \subset \Sigma'$. Assume $\Sigma' \setminus \Sigma_\infty \neq \emptyset$, then there is an $\alpha_n \in \Sigma' \setminus \Sigma_\infty$. Now either $\alpha_n \in \Sigma_n$ or $\neg\alpha_n \in \Sigma_n$. If it is $\alpha_n \in \Sigma_n$, then $\alpha_n \in \Sigma_\infty$, which can't be because of the way α_n was chosen. But if $\neg\alpha_n \in \Sigma_n$, then $\neg\alpha_n \in \Sigma_\infty \subset \Sigma'$, and Σ' would contain both α_n and $\neg\alpha_n$, thus not be consistent. So the set $\Sigma' \setminus \Sigma_\infty = \emptyset$ and thus $\Sigma_\infty = \Sigma'$. Therefore Σ_∞ is a maximal consistent theory.

Let us define $Val : \mathcal{L}_S \to \{0, 1\}$ as such:

$$Val(\alpha) = \begin{cases} 1 & \text{if } \alpha \in \Sigma_\infty \\ 0 & \text{if } \alpha \notin \Sigma_\infty \end{cases}$$

To prove that Val is indeed an (extended) evaluation, we need to show two things: $Val(\bot) = 0$ and $Val(\alpha \to \beta) = Val(\alpha) \to' Val(\beta)$, where $\to$' refers to table 2. The first one is easy, because $Val(\bot) = 0$ if $\bot \notin \Sigma_\infty$, which is the case because Σ_∞ is consistent. For the second one we will look at the three cases:

- $Val(\beta) = 1$: Regardless of the value of $Val(\alpha)$, $Val(\alpha) \to' Val(\beta) = 1$. To have $Val(\alpha \to \beta) = 1$, we need to show that $\alpha \to \beta \in \Sigma_\infty$ when $\beta \in \Sigma_\infty$. From axiom 1, $\beta \to (\alpha \to \beta)$, we can use the deduction theorem 3.11 to show $\beta \vdash \alpha \to \beta$. So when $\beta \in \Sigma_\infty$, we have $\Sigma_\infty \vdash \alpha \to \beta$, which means

[5]As mentioned in footnote 3, if we allow the possibility of an uncountable signature, this proof would require the use of Zorn's Lemma.

$\alpha \to \beta \in \Sigma_\infty$ because of Lemma 3.19. So when $Val(\beta) = 1$, we have $Val(\alpha \to \beta) = Val(\alpha) \to' Val(\beta)$.

- $Val(\alpha) = 0$: From $\{\neg\alpha, \alpha\} \vdash \bot$ and $\bot \vdash \beta$ we can conclude $\neg\alpha \vdash \alpha \to \beta$. Now we follow a similar proof as above. Because $Val(\alpha) = 0$, we have $\alpha \notin \Sigma_\infty$. So, following Lemma 3.19, we know $\neg\alpha \in \Sigma_\infty$. So $\Sigma_\infty \vdash \alpha \to \beta$, so $Val(\alpha \to \beta) = 1 = Val(\alpha) \to' Val(\beta)$.
- $Val(\alpha) = 1$ and $Val(\beta) = 0$: Since now $Val(\alpha) \to' Val(\beta) = 0$, we need to show that $\alpha \to \beta \notin \Sigma_\infty$. If $\alpha \to \beta \in \Sigma_\infty$, then because $\alpha \in \Sigma_\infty$, we have $\Sigma_\infty \vdash \beta$. But $Val(\beta) = 0$, so that can't be the case. It has to be that $\alpha \to \beta \notin \Sigma_\infty$, and thus that $Val(\alpha \to \beta) = 0 = Val(\alpha) \to' Val(\beta)$.

With this, we have shown that $Val(\alpha \to \beta) = Val(\alpha) \to' Val(\beta)$, which means that Val is indeed an evaluation. Now, because $\Sigma \subset \Sigma_\infty$, we have $Val(\Sigma) = 1$. So when Σ is consistent, it has a model. □

With this, we have proven the fundamental theorem of propositional logic.

Corollary 3.20. *The Fundamental Theorem of Propositional Logic implies that propositional logic is sound and complete.*

Proof. Generally sound and generally complete, introduced respectively in lemma 3.15 and 3.16, are stronger properties then sound and complete. This is seen by choosing for Σ the empty set. □

3.5. **Other Possibilities.** In this section, we heavily relied on the exact phrasing of A. Church's system (Definition 3.9). But as mentioned before, using that one was a choice. There are many different systems we could have chosen. But because of the fundamental theorem of propositional logic, we know that whatever choice of system we made, it is impossible to have one capable of proving more true formulas then A. Church's system.

But still, one might ask what a different deduction system on propositional logic would look like. If we keep the semantics the same, and want the final logic to be both sound and complete, we can quickly find some limitations. One thing they all need to have in common, is that their axioms are tautologies. We use this fact in the soundness proof. If it were not the case, then there would be proofs of formulas that were false under certain conditions.

Another way that they might be different is in the distinction of deduction rules and axioms. While treated very differently, we can have these shift around. One may allow deduction rules to not require an input. Then you can use the rule to always generate an output, which makes it effectively act like an axiom.

Furthermore, as was hinted at when the definition of a system was introduced, this is only one of many ways you can define what a formal proof is. Two other options that have their advantages and disadvantages are proof trees and natural deduction [2].

With this, we close the section on propositional logic. It is important that we have laid the groundwork for the different alternate constructions we can extend upon propositional logic.

4. Modal Logic

Before we can truly understand temporal logic, we need to know what modal logic is first. As will soon be explained, temporal logic is one of the four prominent schools

of modal logic. But as is usual when studying modal logic, the general properties will be explored through alethic logic. This leads us to show the soundness and completeness of alethic logic, but also discuss more specific systems in section 4.6. The setup of this section mirrors the previous section, but where mentioned proofs were adapted from other works.

4.1. **What is modal logic.** Propositional logic formalises how general propositions interact with each other. But the general description does not capture more complex statements. For example, let us look at the statement *it is possible for a nine-tailed fox to exist*. From our general knowledge, we can understand this statement without any trouble. We could try to represent it in propositional logic, but we would fail to capture its interaction with related propositions, such as *a nine-tailed fox exists* and *it is necessary for a nine=tailed fox to exist*. How the modifiers *possible* and *necessary* interact with each other goes beyond what propositional logic is designed for. In linguistics, these modifiers of simple propositions are called modals, from which the logic derives its name.

The list of all modals is too broad to enumerate, and like language is ever changing. So modal logic necessarily has a broad scope to encompass many of them. It achieves this via extending the concept of models from propositional logic to a more complex setting. Instead of only dealing with one evaluation function, there is an entire universe of different worlds, each with their own evaluation. This universe is captured in an index set, which I will call T. Not all of the worlds are accessible from each other, which is captured in an accessibility relation on the set T. This relation gives us a way to define how modals should interact with each other.

Unfortunately, a system that would be capable to describe every single model, were it to exist, would be quite useless in its complexity. Instead, different subsets of modals are described somewhat independently of each other. The four most common of these schools are [3]:

- **Alethic Logic**[6], which studies the modals *is possible* and *is necessary*.
- **Deontic Logic**, which studies the modals *is permissible*, *is obligatory* and *is forbidden*.
- **Doxastic Logic**, which studies the modals of the form *X believes that*.
- **Temporal Logic**, which studies the modals *will once be in the future*, *will from now on be*, *has once been in the past* and *has been up till now*.

These four systems have many uses. Alethic Logic is common to find in metaphysics, deontic logic has it usage in theology and law, doxastic logic is prominent in game theory and economics, and temporal logic has already found success in formal specification and verification, i.e. proving that a computer program meets its specification.

Of these four, Alethic logic is the most basic, as its results are used within the other fields. To that end, we will first explore Alethic logic, before we turn to the specifics of temporal logic.[7] This means that in the next section, both concepts specific to alethic logic as well as general modal logic concepts are introduced. Similar to propositional logic, we will first look at the syntax before we discuss the semantics and deduction in turn. Then the soundness and completeness of alethic logic is shown. While the

[6]Many authors refer to alethic logic as modal logic, while still caling the overarching idea modal logic. Alethic logic was chosen to prevent confusion.

[7]For an explanation of the other two main schools, I recommend [4] for deontic logic and [5] for doxastic logic.

general setup of the proofs are similar, the introduction of frames give rise to unique complications.

4.2. **Syntax of Alethic Logic.** The syntax of modal logic extends the syntax of propositional logic. In alethic logic we find two new symbols: $\Box$ and $\Diamond$. These are to be understood as *necessary* and *possible* respectively. We will take $\Box$ to be more fundamental[8], extending the three rules for formula construction of propositional logic with a fourth one:

(4) If φ is a formula, then $\Box\varphi$ is also a formula.

The table of abbreviations will also be extended with an additional item: $\Diamond\varphi$ is an abbreviation of $\neg\Box\neg\varphi$. Both of these symbols will have the same binding strength as $\neg$ when it comes to ordering with brackets. We still use the symbol $\mathcal{L}_S$ to describe the set of all formulas, because there is still a dependency on the signature S. Since modal logic extends propositional logic, it is useful to describe the subset of formulas which could be constructed by the propositional logic syntax with $\mathcal{L}_S^{prop}$. This set of formulas will inherit some properties from propositional logic.

4.3. **Semantics.** First, we have to look at the concept of modal models, before we can look at the specifics of alethic logic semantics. As already mentioned in section 4.1, the models in modal logic are more complicated than for propositional logic. To this end, we make a distinction between a model and a proper model. But before we continue, we need to understand what a frame is in modal logic.

4.3.1. *Frames and Models.*

Definition 4.1. A frame $\mathcal{F} = (T, R)$ consists of an arbitrary non-empty set T called the universe,[9] and an arbitrary (binary) relation R on T called the accessibility relation.

The first thing to notice about the definition of frames is that it encompasses a lot of different constructions. Every partition, every partial order and even sets with an empty relation fit this definition. But this is necessary to make it fit all the different ways modal logic can be expressed. But within each of the four schools mentioned above, each of them will limit the frames that are actually taken in consideration. Unfortunately, there is not much agreement on what these are for alethic logic, as there are arguments for the broad definition 4.1 to already be fitting. A different approach would require R to be at least reflexive, as otherwise one can encounter that an formula φ is true, but the formula "φ is possible", $\Diamond\varphi$ is false. We will return to this question later. The properties of T and R are usually also attributed to the frame itself. So an irreflexive frame is a frame such that R is irreflexive, and a finite frame is a frame such that T is finite.

The set T is usually interpreted as the set of possible worlds, but different schools like temporal logic can have different names.

Each world is going to act like an instance of a propositional model, with an associated $val : S \to \{0, 1\}$. So we can say that at world $x \in T$ the formula $p \to q$ is true, because the val at x has $val(p) = val(q) = 1$. But instead of looking at each

[8]This is an arbitrary choice, which is very similar to the question of whether in firs order logic $\forall$ or $\exists$ ought to be more fundamental.

[9]The symbol T is used for the set because it is the most appropriate symbol in temporal logic. To prevent confusion, it also used here where the symbol will come across as arbitrary.

point individually, the modal evaluation function will be $\pi : T \times S \to \{0, 1\}$, which represents all of the evaluation functions for each world.

4.3.2. *Proper Models and Evaluations.* Although they look similar to those in propositional logic, the function π, which is called a model, is insufficient for the task that models had in propositional logic. We cannot extend a modal model to a useful evaluation function of formulas, since the truth value of $\Box\varphi$ should depend on the accessibility relation R. To remedy this, we introduce proper models.

Definition 4.2. A proper model in modal logic $(\mathcal{F}, \pi)$ consists of a frame $\mathcal{F} = (T, \pi)$ and a model $\pi : T \times S \to \{0, 1\}$.

It is called a proper model because it is the minimal required information to uniquely define an evaluation function for all of $\mathcal{L}_S$. The exact way depends of the new symbols introduced in the syntax, which means we return to the specifics of alethic logic.

Lemma 4.3. *A proper model $(\mathcal{F}, \pi)$ in alethic logic defines a unique evaluation function $\Phi : T \times \mathcal{L}_S \to \{0, 1\}$, as follows: for each $x \in T$:*

$$\begin{aligned}
\Phi(x, p) &= \pi(x, p) \quad \textit{for each } p \in S \\
\Phi(x, \bot) &= 0 \\
\Phi(x, \varphi \to \psi) &= \Phi(x, \varphi) \to' \Phi(x, \psi) \\
\Phi(x, \Box\varphi) &= 1 \quad \textit{if for each } y \in T \textit{ with } xRy : \Phi(y, \varphi) = 1 \\
\Phi(x, \Box\varphi) &= 0 \quad \textit{if there exists an } y \in T \textit{ with } xRy : \Phi(y, \varphi) = 0
\end{aligned}$$

Where the symbol $\to'$ refers to the binary operator shown in table 2.

Proof. By noting that the last two lines are exhaustive, it is trivial to see that the construction outlined in the lemma indeed gives us a function $\Phi : T \times \mathcal{L}_S \to \{0, 1\}$. So we only need to busy ourselves with the uniqueness of Φ. If we limit the function π to one point $x \in T$, the first three definitions of Φ mirror those of lemma 3.3. So were there two possible extensions Φ and Φ' fulfilling the definitions, then they can't differ from evaluations in a formula that doesn't contain the $\Box$ symbol. Otherwise we could find a counterexample to lemma 3.3.

So from the lemma, we can conclude that for each $t \in T$, there exist a unique map $val : \mathcal{L}_S^{prop} \to \{0, 1\}$ where $\mathcal{L}_S^{prop}$ is the set of all formulas that does not contain the $\Box$ symbol. These evaluations can be combined into one map $\Phi^* : T \times \mathcal{L}_S^{prop} \to \{0, 1\}$, which is uniquely determined by π.

Now, let there be two evaluation function Φ and $\hat{\Phi}$ that fulfil the definition, and let us look at the formula $\Box\varphi$ where $\varphi \in \mathcal{L}_S^{prop}$. Since Φ^* is unique for a given π, it must be the case that for each $x \in T$: $\Phi(x, \varphi) = \Phi^*(x, \varphi) = \hat{\Phi}(x, \varphi)$. Now $\Phi(x, \Box\varphi) = 1$ means that for every $y \in T$ with xRy, $\Phi(y, \varphi) = 1$, which requires that for every $y \in T$ with xRy, $\hat{\Phi}(y, \varphi) = 1$, and thus that $\hat{\Phi}(x, \Box\varphi) = 1$. Similar, when $\Phi(x, \Box\varphi) = 0$, there exists an $z \in T$ with xRz where $\Phi(z, \varphi) = 0$, for which also $\hat{\Phi}(z, \varphi) = 0$, which requires that $\hat{\Phi}(x, \Box\varphi) = 0$. So $\Phi(x, \Box\varphi) = \hat{\Phi}(x, \Box\varphi)$.

This argument can be repeated indefinitely, and since each formula is finite, this will be enough to show that $\Phi = \hat{\Phi}$. □

It is clear why we need proper models, as Φ is undefined if we have not chosen a relation R. This dependence also means that the implicit function interpretation

$\varphi \mapsto \varphi'$ from propositional logic is insufficient. Depending on R, the same φ' with the same $\pi(x, p_i)$ can evaluate to either 0 or 1. This can be solved if the mapping $\varphi \mapsto \varphi'$ depends on the frame $\mathcal{F}$, but not on the model π. Since this removes the ambiguity, it is a unique mapping. The notation does need to be updated: Let $\mathcal{L}_{S,n}$ still be the set of formulas with n atomic propositions. For a fixed $\mathcal{F}$, we have the mappings $\mathcal{L}_{S,n} \to \{T \times \underline{2}^n \to \underline{2}\}$, denoted by $\varphi \mapsto \varphi'_{\mathcal{F}}$, such that $\Phi(x, \varphi) = \varphi'_{\mathcal{F}}(x, \pi(x, p_1), \ldots, \pi(x, p_1))$.

The reason that we distinguish between proper and ordinary models is clarity. It will be more natural to be talking about the two parts that make a proper model separately. The part that behaves similarly to models in propositional logic tends to be more flexible, while the frames are very rigid, since they represent an underlying structure of the universe T. This distinction comes forward when we try to apply the propositional logic concept of tautologies to modal logic.

4.3.3. *Modal Tautologies.* In propositional logic, a tautology was a formula whose truth value was model-independently true (Definition 3.4). This can be extended to modal logic.

Definition 4.4. A formula $\varphi \in \mathcal{L}_S$ is a modal tautology if for every frame $\mathcal{F}$ and every evaluation π on $\mathcal{F}$ and each world $x \in T$, $\Phi(x, \varphi) = 1$. This is written as $\models \varphi$.

Being a modal tautology sounds like a far stronger property than being a tautology in propositional logic. And in some sense it is. In propositional logic, a formula only needs to be true for 2^n different evaluations. But because there is no limit on the size of T, there is an infinite number of possible frames, each of which can have a large or infinite number of possible $\pi : T \times S \to \{0, 1\}$. So it is indeed the case that being a modal tautology requires a much stronger proof, as finite proofs by exhaustion won't be guaranteed. But as it turns out, there are more formulas that are modal tautologies then there are tautologies in propositional logic. This is due to there being more formulas in total, in combination with the following proposition.

Proposition 4.5. *Each formula $\varphi \in \mathcal{L}_S^{prop}$ that is a tautology in propositional logic is also a modal tautology: $\models \varphi$.*

Proof. Since φ is a tautology in propositional logic, for each evaluation val, we have $Val(\varphi) = 1$. Now, for a given frame $\mathcal{F}$, a modal evaluation π restricted to world $x \in T$ and to the formulas $\mathcal{L}_S^{prop}$ must be an evaluation val. And restricting the extension of π, Φ to world $x \in T$ and to the formulas $\mathcal{L}_S^{prop}$ gives us an extended evaluation Val. Since the extensions $\pi \mapsto \Phi$ and $val \mapsto Val$ are unique, they must match up. So $\Phi(x, \varphi) = Val(\varphi) = 1$. This is true regardless of which frame we chose, so for every frame $\mathcal{F}$ and every evaluation π, at each $x \in T$, $\Phi(x, \varphi) = 1$. Thus $\models \varphi$. □

Intuitively this makes sense, since each world was envisioned as having its own propositional logic internally. So at each world $x \in T$, a propositional tautology must hold. And since the interaction between different worlds does not come into play when we look at formulas from $\mathcal{L}_S^{prop}$, they must be tautologies regardless of the size of T or of the nature of the relation R. With the next lemma, it is possible to extend this to even more modal tautologies.

Lemma 4.6. *If $\models \varphi$, where φ has n atomic propositions, and $\{\alpha_1, \ldots \alpha_n\} \subset \mathcal{L}_S$, then:*

(1) $\models \Box\varphi$, *and*
(2) $\models \varphi^*$, *where* φ^* *is obtained by replacing the instance of* p_i *in* φ *with* α_i.

Proof. (1) Since $\models \varphi$, for any give frame $\mathcal{F}$ and any evaluation π, for all $y \in T$: $\Phi(y, \varphi) = 1$. Thus for all y with xRy it is also the case that $\Phi(y, \varphi) = 1$, so $\Phi(x, \Box\varphi) = 1$. Since this is independent of the frame, the evaluation or the specific world, $\models \Box\varphi$.

(2) If φ^* was not an tautology, then there is a frame $\mathcal{F}$, an evaluation π^* and a world $x \in T$ such that $\Phi^*(x, \varphi^*) = 0$. Now define π, such that $\pi(x, p_i) = \Phi^*(x, \alpha_i)$. Then

$$\begin{aligned}\Phi(x, \varphi) &= \varphi'_{\mathcal{F}}(x, \pi(x, p_1), \ldots, \pi(x, p_n)) \\ &= \varphi'_{\mathcal{F}}(x, \Phi^*(x, \alpha_1), \ldots, \Phi^*(x, \alpha_n)) = \Phi^*(x, \varphi^*) = 0.\end{aligned}$$

So then there is a frame $\mathcal{F}$, with an evaluation π such that at a world $x \in T$, $\Phi(x, \varphi) = 0$. This contradicts $\models \varphi$. So $\models \varphi^*$. □

With the use of Lemma 4.6, we can find a whole slew of tautologies which originate from propositional logic. But the modal tautologies which have no connection to propositional tautologies are more interesting. Let us look at an example.

Example 4.7. An important modal tautology in alethic logic is

$$\models \Box(\alpha \to \beta) \to (\Box\alpha \to \Box\beta).$$

To show it is indeed an modal tautology, assume the contrary. Then there is a frame $\mathcal{F}$ and an evaluation π such that there is an $x \in T$ where

$$\Phi(x, \Box(\alpha \to \beta) \to (\Box\alpha \to \Box\beta)) = 0.$$

It follows from the definition of $\to'$ that this is only the case when

$$\Phi(x, \Box(\alpha \to \beta)) = 1, \Phi(x, \Box\alpha) = 1 \text{ and } \Phi(x, \Box\beta) = 0.$$

From the first and second requirement, it must be the case that for each xRy, $\Phi(y, \alpha \to \beta) = 1$ and $\Phi(y, \alpha) = 1$. This would require that for each xRy, $\Phi(y, \beta) = 1$. This means that $\Phi(x, \Box\beta) = 1$, contradicting the third requirement. Thus it must be the case that for each frame and each model

$$\Phi(x, \Box(\alpha \to \beta) \to (\Box\alpha \to \Box\beta)) = 1.$$

So the formula is a modal tautology.

4.3.4. *Frame Dependencies and Definability.* As will be shown in section 4.5, this definition of modal tautology is adequate in the sense that it can form the counterpart to a sound and complete deduction system for alethic logic[10]. But that system will not be as expressive as one might want. For example, let us look at the formula $p \to \Diamond p$ we encountered earlier. If we interpreted alethic logic as how the words necessary and possible are used in language, then one property could be that if p is true, then possible p must be true, since we have an example where it is actually true, so it must be possible. So we would expect that $p \to \Diamond p$ be true at every $x \in T$, regardless of what frame we choose. But one can easily find counterexamples where $p \to \Diamond p$ does not hold. The most trivial case, R being an empty relation, would suffice. So $p \to \Diamond p$ is not a modal tautology. In propositional logic, we extended the definition

[10]It is also adequate for the other schools of modal logic, but only temporal logic is proven in theorem 5.4 below.

of $\models$ to allow for some restrictions on the models that are taken into consideration. We can do the same for modal logic, but there is a more modal concept to capture the properties of formulas like $p \to \Diamond p$.[11]

Definition 4.8. A modal logic formula φ that is model-independently true for a given frame $\mathcal{F}$, is called a modal tautology with respect to $\mathcal{F}$. We also say that φ is necessarily true in $\mathcal{F}$. This is denoted by $\mathcal{F} \models \varphi$.

So, if $T = \mathbb{N}$, and R is the normal order $\leq$, then $(T, R) \models p \to \Diamond p$. Now, being necessarily true in a frame is a weaker property then being a modal tautology. Nonetheless, the rules for constructing necessarily true formulas from smaller ones are similar.

Lemma 4.9. *Let $\mathcal{F}$ be a frame, $\varphi \in \mathcal{L}_S$ and $\{\alpha_1, \dots \alpha_n\} \subset \mathcal{L}_S$. Then the following hold:*

(1) *If $\models \varphi$, then $\mathcal{F} \models \varphi$.*
(2) *If $\mathcal{F} \models \varphi$, then $\mathcal{F} \models \Box\varphi$.*
(3) *If $\mathcal{F} \models \varphi$, then $\mathcal{F} \models \varphi^*$, where φ^* is obtained by replacing the instance of p_i in φ with α_i.*

Proof. For (1), if for each frame $\mathcal{F}$ for every evaluation π at each $x \in T : \Phi(x, \varphi) = 1$, then this is also the case for a specific $\mathcal{F}$. The proofs for (2) and (3) are the same as the proof for Lemma 4.6, where instead of talking about any frame $\mathcal{F}$, we talk about the specific frame $\mathcal{F}$. □

This can then be generalised to classes of frames.

Definition 4.10. Let C be a class of frames. We write $\mathsf{C} \models \varphi$ if for each $\mathcal{F} \in \mathsf{C} : \mathcal{F} \models \varphi$.

Now we can return to the formula $p \to \Diamond p$. Let

$$\mathsf{R} = \{\mathcal{F} | \mathcal{F} \text{ is a frame and } \mathcal{F} \models p \to \Diamond p\}.$$

Now we can ask ourselves what can we say about frames that are in R. Because the formula $p \to \Diamond p$ will be necessarily true, we have restrictions on R. For every $x \in T$, we can define the evaluation π_x such that $\pi_x(y, p) = \delta_{xy}$. Since this is a valid evaluation, in $\mathcal{F} \in \mathsf{R} : \Phi_x(x, p \to \Diamond p) = 1$. Since per definition $\Phi_x(x, p) = 1$, we have $\Phi_x(x, \Diamond p) = 1$. So there must exist an $y \in T$ with xRy and $\Phi_x(y, p) = 1$. But the only candidate world is x, since everywhere else, $\Phi_x(y, p) = 0$. Thus it must be the case that xRx, i.e. $\mathcal{F}$ is reflexive.

And when we look in the opposite direction, it is trivial to show that any reflexive frame $\mathcal{F}$ has $\mathcal{F} \models p \to \Diamond p$, and thus $\mathcal{F} \in \mathsf{R}$. This means that there is a correlation between a frame being reflexive and the formula $p \to \Diamond p$ being necessarily true in $\mathcal{F}$. This can be generalised to the notion of definability.

Definition 4.11. A formula φ defines a class of frames C (within a class K) if for every frame $\mathcal{T}$ (in K), $\mathcal{T} \models \varphi$ if and only if $\mathcal{T}$ belongs to C.

A class can be just any arbitrary set of frames, but most of the time it is defined by its members sharing a specific property. A formula φ that defines a class is called the defining formula for that class. This is not a unique property. If a defining formula

[11] While a modal concept of limiting the models can sometimes be used, it will not be necessary for us.

φ is tautologically equivalent to another formula, i.e. $\models \varphi \leftrightarrow \psi$, then ψ is also a defining formula for the same class.

A defining formula acts as an indicator for certain properties of a frame. If we are given a frame $\mathcal{F}$, and we can prove that $\mathcal{F} \models \varphi$, then we have also proven that $\mathcal{F}$ has the property defined by φ. Technically, any formula defines a class. But for most of them it will be an interesting class. Most formulas do not behave consistently enough for there to be any frame in which the formula is necessarily true. Consequently, these formulas the empty class of frames. Meanwhile, every modal tautology defines the entire class of frames. So, formulas which are interesting need to lay between these two extremes when it comes to consistency.

In the definition, the parts between brackets come into play if we are already looking within a smaller class of frames. While sometimes avoidable, such as described in Lemma 4.12, dealing with smaller classes is required when we want to look at a property such as irreflexivity. Irreflexivity is a valid property for frames to have, but it can be shown that there exist no formula which defines the class of irreflexive frames. So, while it is possible to define the class of dense partial orders within the class of partial orders, since there is no formula which defines the class of partial order, there is no formula which directly defines the class of dense partial orders either.

The main way one can avoid having to deal with formulas that define a class within a class is by applying the following Lemma.

Lemma 4.12. *If the class* C *is defined by* φ*, and the class* D *within the class* C *is defined by* ψ*, then the class* D *is directly defined by the formula* $\varphi \wedge \psi$*.*

Proof. Let E be the class defined by the formula $\varphi \wedge \psi$. Take $\mathcal{F} \in \mathsf{D}$. Since $\mathsf{D} \subset \mathsf{C}$, both $\mathcal{F} \models \varphi$ and $\mathcal{F} \models \psi$, so $\mathcal{F} \models \varphi \wedge \psi$. Thus $\mathcal{F} \in \mathsf{E}$. Now take $\mathcal{G} \in \mathsf{E}$. Then $\mathcal{G} \models \varphi \wedge \psi$. Since $\mathcal{G} \models \varphi$, $\mathcal{G} \in \mathsf{C}$. And now that $\mathcal{G} \in \mathsf{C}$ and $\mathcal{G} \models \psi$, it follows that $\mathcal{G} \in \mathsf{E}$. Thus $\mathsf{E} = \mathsf{D}$, so D is defined by $\psi \wedge \varphi$. □

4.4. **Deduction.** Now we turn our attention to deduction. As already alluded to in section 3.3, the theory of deduction will be more similar to what was described in propositional logic. Unlike last section, there are only a few more details that need be introduced, which will play an important role in section 4.6. The general notation introduced here applies to all of modal logic, while the examples are specific to alethic logic.

4.4.1. *The Logic of Modal Logic.* The main difference between deduction in propositional logic and deduction in modal logic, is the variety of systems which are studied. This is due to the fact that the semantics of propositional logic is unambiguous, while it has already been touched upon in the last subsection that there is no unanimity on what the properties of proper alethic frames are. And this has an impact on the deduction side of modal logic. So we will introduce some notation on different systems, and prove some preliminary results which will be used in the completeness proof.

What a deduction system is remains the same in modal logic as it was in propositional logic (definition 3.7). But the notation of $\vdash$ as it was introduced in propositional logic has some issues. We can still use the meaning of $\vdash \varphi$, but the notion of deriving from a set of assumptions, $\Sigma \vdash \varphi$, will give us problems in this new context. Were we to take the same definition, the statement $\alpha \vdash \Box\alpha$ would be trivially true once Kripke's logic is introduced. But the deduction theorem would then imply that

$\vdash \alpha \to \Box\alpha$, which is clearly not a modal tautology. So while one approach to this issue is abandoning the deduction theorem, better results are obtained if the definition is slightly altered.

Definition 4.13. A formula φ can be derived from a theory Σ if there exists a formula $\boldsymbol{\alpha}$ that is constructed by a conjunction of a finite number of formulas from Σ such that $\vdash \boldsymbol{\alpha} \to \varphi$. This is denoted by $\Sigma \vdash \varphi$.

With this change, we only need to discuss ambiguity. In propositional logic, the symbol $\vdash$ was unambiguous, as only the A. Church's system (definition 3.9) was ever considered. But this is not the case in modal logic. So whenever we have a formal proof of φ within a system $\mathbf{L} = (A_{\mathbf{L}}, D_{\mathbf{L}})$, we will note it as $\vdash_{\mathbf{L}} \varphi$. If there is no system mentioned, it is assumed to be Kripke's system, the most basic system in alethic logic.

Definition 4.14 (Kripke's system **K**)**.** **K** is a system that expands the A. Church's system with one additional deduction rule and one additional axiom scheme.
Additional deduction rule:

Generalisation: As input we take one formula of the form α and we output $\Box\alpha$.
Axiom scheme:

(4) $\Box(\alpha \to \beta) \to (\Box\alpha \to \Box\beta)$.

Axiom scheme 4, also known as distribution axiom, regulates the $\Box$ symbol. Note that an axiom obtained form any of the four schemes is a modal tautology.

K, being the most basic system, will turn out to be sound and complete. But it is not the only system of interest, as we do not always want to describe the entirety of alethic logic. Let us once again return to $p \to \Diamond p$. This formula was enough to define the property of reflexivity, so if we want to find a logic which works nicely with the reflexive framws, extending **K** with the axiom $p \to \Diamond p$ seems like a good candidate. This method of extending a system can be generalised to the following addition rule.

Definition 4.15. A system **L** can be extended by a set of additional axioms A to form a new logic. This new system is called $\mathbf{L}+A$. Further, $\mathbf{L}+\{\varphi\}$ is an abbreviation of $\mathbf{L}+\varphi$, and $\mathbf{L}+A\cup B$ is written as $\mathbf{L}+A+B$.

While the systems $\mathbf{L}+\varphi$ and $\mathbf{L}+\psi$ might look different, they can still be equivalent in their effectiveness. Two systems are equivalent if the set of formulas they can prove are equal. This leads to the following lemma.

Lemma 4.16. *Let* **L** *be a system, φ a formula and A and B finite sets of formulas. Then:*

1. **L** *and* $\mathbf{L}+\varphi$ *are equivalent if* $\vdash_{\mathbf{L}} \varphi$.
2. $\mathbf{L}+A$ *and* $\mathbf{L}+B$ *are equivalent, if for every* $\beta \in B$, $\vdash_{\mathbf{L}+A} \beta$ *and for every* $\alpha \in A$, $\vdash_{\mathbf{L}+B} \alpha$.

Proof. For (1), assume $\vdash_{\mathbf{L}} \varphi$. Let $\vdash_{\mathbf{L}+\varphi} \alpha$. Then in the proof, we can replace the step where φ is evoked with the proof of $\vdash_{\mathbf{L}} \varphi$. Thus $\vdash_{\mathbf{L}} \alpha$. And trivially $\vdash_{\mathbf{L}} \alpha$ implies that $\vdash_{\mathbf{L}+\varphi} \alpha$.

For (2), since for every $\beta \in B$, $\vdash_{\mathbf{L}+A} \beta$ we can repeatedly use (1) to show that $\mathbf{L}+A$ is equivalent to $\mathbf{L}+A+B$. Similarly, from $\alpha \in A$, $\vdash_{\mathbf{L}+B} \alpha$ we can conclude that $\mathbf{L}+B$ is equivalent to $\mathbf{L}+A+B$. Since equivalence is clearly a transitive property, $\mathbf{L}+A$ is equivalent to $\mathbf{L}+B$. □

We will return to the properties of systems of the form $\mathbf{K}+A$ in section 4.6.

4.4.2. *Modal Deduction Theorem.* For the completeness proof we will again be dealing with consistent and inconsistent theories. But we cannot just use the results from propositional logic, since it was built on the deduction theorem for a different definition of $\vdash$. So we will first have to give a new prove for the deduction theorem.

Lemma 4.17. *If* **L** *is a system that can deduce all propositional tautologies, then* $\Sigma \cup \{\varphi\} \vdash \psi \Leftrightarrow \Sigma \vdash \varphi \to \psi$.

Proof. Per definition $\Sigma \vdash \varphi \to \psi \Leftrightarrow \vdash \boldsymbol{\alpha} \to (\varphi \to \psi)$. Since the propositional formula $\gamma \to (\alpha \to \beta) \leftrightarrow \gamma \wedge \alpha \to \beta$ is a propositional tautology, we can apply it for $\gamma = \boldsymbol{\alpha}$, $\alpha = \varphi$ and $\beta = \psi$. This gives us that

$$\vdash \boldsymbol{\alpha} \to (\varphi \to \psi) \Leftrightarrow \vdash \boldsymbol{\alpha} \wedge \varphi \to \psi \Leftrightarrow \Sigma \cup \{\varphi\} \vdash \psi.$$

□

With the recovered deduction theorem, we can reuse the proofs from propositional logic to gain the same results for modal logic. But we again need to clear any ambiguity when it comes to which system we are discussing. An **L**-inconsistent theory Σ is still a theory such that $\Sigma \vdash_{\mathbf{L}} \bot$. Then from lemma 3.12 we can still conclude that if $\nvdash_{\mathbf{L}} \varphi$, then the set $\{\neg\varphi\}$ is **L**-consistent.

4.5. **Soundness and Completeness.** With all the prerequisites out of the way, we will prove the soundness and completeness of alethic logic. We will here only focus on Kripke's system, and in the next section look on finding sound and complete systems for different classes of frames. Unlike in propositional logic, where a more general theorem was proven, in modal logic we will only look at simple soundness and completeness. The more general case is not as important for modal logic, since instead we will be looking at different systems in section 4.6.

Theorem 4.18. *Modal logic is sound and complete, i.e.* $\vdash_{\mathbf{K}} \varphi$ *if and only if* $\models \varphi$.

Since we will only be talking about Kripke's system, we will forgo the **K** label for this section. So we will be talking about consistent theories, and use the $\vdash$ symbol without index. The proof will consist of two parts. The proof of soundness will be similar to the proof of Lemma 3.15, as we will again use induction on the length of the proof for the formula φ. On the other hand, the proof of completeness in propositional logic relied on the semantic deduction theorem, which has no counterpart in modal logic.

4.5.1. *Soundness.*

Lemma 4.19. *The system* K *is sound, i.e.* $\vdash \varphi \Rightarrow \models \varphi$.

Proof. The proof is exactly the same as that of Lemma 3.15, except for two places. First, a change of terminology is necessary. Instead of looking at val and its extension Val, we will be looking at each frame $\mathcal{F}$ at each world $x \in T$ at π and its extension Φ. Secondly, it is now also possible that the last formula in a proof was gotten via generalisation. This means that it can be the case that $\varphi = \Box\psi$, where ψ appears earlier in the proof. Since $\vdash \psi$ has a shorter proof, we use the induction hypothesis to conclude that $\models \psi$. Then we use Lemma 4.6 to conclude that $\models \Box\psi$. □

Note that, as long as the set A only consist of tautologies, this proof holds. This will come into play in the next section.

4.5.2. *Completeness.*

Lemma 4.20. *The system* K *is complete, i.e.* $\models \varphi \Rightarrow \vdash \varphi$.

The proof is adapted from J. van Benthem, *Modal Logic for Open Minds* [6], and will run through multiple steps. We will construct what is known as the canonical model, which is a proper model that has some nice properties. In this proper model, the contrapositive of Lemma 4.20 can be easily shown. What has to be said before the proof begins, is that it relies heavily on the interplay between $\Box$ and $\Diamond$, which deviates from the proofs up till now, where only $\Box$ was discussed. At one point we will need to take $\Diamond$ as more fundamental in an induction proof. But by diligently applying $\Box\varphi = \neg\Diamond\neg\varphi$, no difficulties should arise.

Before we can go and prove Lemma 4.20, we need to establish some properties of maximal consistent theories.

Lemma 4.21. *In modal logic, a maximally consistent theory* Σ *has the following properties:*

(1) $\Sigma \vdash \varphi$ *if and only if* $\varphi \in \Sigma$.
(2) $\varphi \notin \Sigma$ *if and only if* $\neg\varphi \in \Sigma$.

Proof. For (1), assume $\Sigma \vdash \varphi$ but $\varphi \notin \Sigma$. Since Σ is maximal, the set $\Sigma \cup \{\varphi\}$ must be inconsistent. Using the deduction theorem 4.17, $\Sigma \cup \{\varphi\} \vdash \bot$ implies $\Sigma \vdash \varphi \to \bot$. But combining that with $\Sigma \vdash \varphi$ would mean that $\Sigma \vdash \bot$, which cannot be the case. Thus $\varphi \in \Sigma$. The other direction is trivial.

For (2), assume $\varphi \notin \Sigma$. Then $\Sigma \cup \{\varphi\} \vdash \bot$. Using the deduction theorem again, we conclude that $\Sigma \vdash \varphi \to \bot$. Using (1), we can then conclude that $\neg\varphi \in \Sigma$. The other direction is again trivial, for the inclusion of $\neg\varphi$ excluded φ in any consistent theory. □

We can now start constructing the canonical (proper) model $(\mathcal{S}, R_c, \pi_c)$, also known as the Henkin (proper) model. It is a proper model on a very special set, namely the class of all maximally consistent sets $\mathcal{S}$. On this set, we construct quite a complicated accessibility relation R_c, namely

$$\Sigma R_c \Delta \text{ if and only if } \forall_{\alpha\in\Delta}(\Diamond\alpha \in \Sigma).$$

Each world in $\mathcal{S}$ is a set of formulas, so we can have π_c be defined as $\pi_c(\Sigma, p) = \chi_\Sigma(p)$.

The canonical model has a very nice feature, "Truth coincides with membership." But to prove this, we first need some properties of the accessibility relation R_c.

Lemma 4.22. *Let* Σ *be a consistent set. Then* $\Diamond\varphi \in \Sigma$ *if and only if* $\exists_{\Delta\in\mathcal{S}}\, \Sigma R_c \Delta$ *and* $\varphi \in \Delta$.

Proof. The only if part is trivial, as $\Diamond\varphi \in \Sigma$ is required for each $\varphi \in \Delta$.

For the if part, we need to construct a maximally consistent set which satisfies the requirements. First, let $\Gamma = \{\alpha | \Box\alpha \in \Sigma\}$. We claim that since $\Diamond\varphi \in \Sigma$, $\Gamma \cup \{\varphi\}$ is consistent. Assume it is not: $\Gamma \cup \{\varphi\} \vdash \bot$. Using the deduction theorem, this can be rewritten as $\vdash \boldsymbol{\alpha} \to \neg\varphi$. By using the generalisation rule and the distribution axiom, we can conclude that $\vdash \Box\boldsymbol{\alpha} \to \Box\neg\varphi$.

With this result, we turn our attention to the set Σ. By construction of Γ, for every $\alpha \in \Gamma$, we have $\Box\alpha \in \Sigma$. By using the factorisation rule for $\Box$,

$$\vdash \Box\alpha \wedge \Box\beta \to \Box(\alpha \wedge \beta),$$

we can conclude that $\Sigma \vdash \Box\boldsymbol{\alpha}$. Showing the correctness of the factorisation rule is left as an exercise to the reader. Combining this result with the previous one, we can conclude that $\Sigma \vdash \Box\neg\varphi$. But since $\Box\neg\varphi = \neg\Diamond\varphi$, we have that both $\neg\Diamond\varphi$ and $\Diamond\varphi \in \Sigma$. This contradiction means that our assumption that Γ was inconsistent was wrong.

Now that we know that $\Gamma \cup \{\varphi\}$ is consistent, we can extend it to a maximally consistent set Λ. The exact method can be the same as the one described in the proof of lemma 3.16, but these details are not relevant. It is clear that $\varphi \in \Lambda$, so we only need to show that $\Sigma R_c \Lambda$. Let $\alpha \in \Lambda$ and suppose that $\Diamond\alpha \notin \Sigma$. Because of lemma 4.21, this implies that $\neg\Diamond\alpha = \Box\neg\alpha \in \Sigma$. This would mean that $\neg\alpha \in \Gamma \subset \Lambda$. But because $\alpha \in \Lambda$, this would make Λ inconsistent. Therefore, $\Diamond\alpha \in \Sigma$ and thus $\Sigma R_c \Lambda$. □

Now we can prove that the canonical modal has the "Truth coincides with membership" property:

Lemma 4.23. *For any $\Sigma \in \mathcal{S}$ and formula φ, in the canonical model $(\mathcal{S}, R_c, \pi_c)$ we have*

$$\Phi_c(\Sigma, \varphi) = 1 \Leftrightarrow \varphi \in \Sigma.$$

Proof. We use induction on the length of the formula φ. The length is defined by the number of occurrences of the $\rightarrow$ and $\Diamond$ symbols. Here we have to apply $\Box = \neg\Diamond\neg$.

If the length is 0, then either $\varphi = \bot$, or $\varphi = p \in S$. If it is the first, then both $\varphi \notin \Sigma$ and $\Phi_c(\Sigma, \varphi) = 0$. If it is the latter, then $\Phi_c(\Sigma, \varphi) = \pi(\Sigma, p) = \chi_\Sigma(p) \Leftrightarrow p \in \Sigma$.

Let this be the case for length n and less, and let φ be of length $n+1$. There are two options, either $\varphi = \alpha_1 \rightarrow \alpha_2$, such that length α_1 plus length α_2 is n, or $\varphi = \Diamond\psi$, such that the length of ψ is n.

If it is the former, then we need to prove the two directions separately. Assume $\Phi_c(\Sigma, \varphi) = \Phi_c(\Sigma, \alpha_1) \rightarrow' \Phi_c(\Sigma, \alpha_2) = 1$. If $\Phi_c(\Sigma, \alpha_2) = 1$, then by the induction assumption $\alpha_2 \in \Sigma$. Since $\alpha_2 \vdash \alpha_1 \rightarrow \alpha_2$ and lemma 4.21 we can conclude that $\alpha_1 \rightarrow \alpha_2 = \varphi \in \Sigma$.

If $\Phi_c(\Sigma, \alpha_2) = 0$, then $\alpha_2 \notin \Sigma$ and according to lemma 4.21 this would mean that $\neg\alpha_2 \in \Sigma$. It also requires that $\Phi_c(\Sigma, \alpha_2) = 0$ such that $\Phi_c(\Sigma, \varphi) = 1$. This would imply that $\neg\alpha_1 \in \Sigma$. Then from $\{\neg\alpha_1, \neg\alpha_2\} \vdash \alpha_1 \rightarrow \alpha_2$ it would be the case that $\alpha_1 \rightarrow \alpha_2 = \varphi \in \Sigma$.

Now assume that $\varphi = (\alpha_1 \rightarrow \alpha_2) \in \Sigma$. It cannot be the case that $\Phi_c(\Sigma, \alpha_1) = 1$ while $\Phi_c(\Sigma, \alpha_2) = 0$, since that would apply that both α_1 and $\alpha_1 \rightarrow \alpha_2 \in \Sigma$, while $\neg\alpha_2 \in \Sigma$. That would make Σ inconsistent.

This proves the case that $\varphi = \alpha_1 \rightarrow \alpha_2$. If, $\varphi = \Diamond\psi$, the proof follows directly form Lemma 4.22.

$$\begin{aligned}\Phi_c(\Sigma, \varphi) = \Phi_c(\Sigma, \Diamond\psi) = 1 &\Leftrightarrow \exists_{\Delta \in \mathcal{S}}\, \Sigma R_c \Delta : \Phi_c(\Delta, \psi) = 1 \\ &\Leftrightarrow \exists_{\Delta \in \mathcal{S}}\, \Sigma R_c \Delta \text{ and } \psi \in \Delta \Leftrightarrow \Diamond\psi \in \Sigma.\end{aligned}$$

In the second to last step, we used the induction assumption, and in the last step we used the lemma. □

Having shown the "Truth coincides with membership" property of the canonical proper model $(\mathcal{S}, R_c, \pi_c)$, we can finally prove the completeness half.

Proof of lemma 4.20. We will show the contrapositive. Let $\nvdash \varphi$. Then from the explanation at the end of section 4.4, we know that $\{\neg\varphi\}$ is consistent. This can be

extended to a maximally consistent set Σ. Then using lemma 4.23, we can conclude from $\varphi \notin \Sigma$ that in the canonical model $(\mathcal{S}, R_c, \pi_c)$ we have $\Phi_c(\Sigma, \varphi) = 0$. Thus we have found a frame with an evaluation such that there is a world where φ is false. Therefore, $\not\models \varphi$. □

4.6. **Axiomatization.** In the previous section, we have proven that the system $\mathbf{K}$ is sound and complete. But section 4.4 had introduced more than just the system $\mathbf{K}$. We can have different systems, such as $\mathbf{K} + \varphi$, and we would like to prove is these are also sound and complete. At first, this seems like a trivial question. Unless φ was already a modal tautology, it is the case that $\vdash_{\mathbf{K}+\varphi} \varphi$, while $\not\models \varphi$. But if the formula φ is a modal tautology, then it was already the case that $\vdash_{\mathbf{K}} \varphi$, meaning that $\mathbf{K} + \varphi$ is equivalent to $\mathbf{K}$. So $\mathbf{K} + \varphi$ is either equivalent to $\mathbf{K}$ and thus not interesting, or it is not sound and thus not interesting.

4.6.1. *Soundness and Completeness to a Class.* While the question of whether $\mathbf{K} + \varphi$ is sound or complete is moot, sometimes the formulas that are deducible by the system $\mathbf{K} + \varphi$ are all necessarily true in some frames. This can be formalised if we introduce a more nuanced definition of being sound and complete.

Definition 4.24. A system $\mathbf{L}$ is sound and complete to a class C if for every $\varphi \in \mathcal{L}_S$:

$$\mathsf{C} \models \varphi \Leftrightarrow \vdash_{\mathbf{L}} \varphi.$$

When we are given a class of frames, for example the class of reflexive frames R, we might ask ourselves which systems are sound and complete with respect to R, or whether one even exists. This is the axiomatization of the class R. But if we already have a system, for example $\mathbf{K} + \varphi$, we would want to know to what class it is sound and complete. Unfortunately, the set C is generally not unique, though the argument proving this is a bit difficult[12]. To fix this issue, we need to introduce some properties of definition 4.24.

Lemma 4.25. *Let* A, B *be classes w.r.t. which the system* $\mathbf{L}$ *is sound and complete. Then* $\mathbf{L}$ *is sound and complete to* $\mathsf{A} \cup \mathsf{B}$.

Proof. We will prove the contrapositive: for for every $\varphi \in \mathcal{L}_S$:

$$\mathsf{A} \cup \mathsf{B} \not\models \varphi \Leftrightarrow \nvdash_{\mathbf{L}} \varphi.$$

Let φ be any formula.

If $\mathsf{A} \cup \mathsf{B} \not\models \varphi$ then there exists a $\mathcal{F} \in \mathsf{A} \cup \mathsf{B}$ such that $\mathcal{F} \not\models \varphi$. Without loss of generality, assume $\mathcal{F} \in \mathsf{A}$. Then $\mathcal{F} \not\models \varphi$ implies that $\mathsf{A} \not\models \varphi$. Now, because $\mathbf{L}$ is sound and complete w.r.t. A, this means that $\nvdash_{\mathbf{L}} \varphi$.

If $\nvdash_{\mathbf{L}} \varphi$, because $\mathbf{L}$ is sound and complete to A, it must be the case that $\mathsf{A} \not\models \varphi$. And because $\mathsf{A} \subset \mathsf{A} \cup \mathsf{B}$, $\mathsf{A} \cup \mathsf{B} \not\models \varphi$, and we have $\mathsf{A} \cup \mathsf{B} \not\models \varphi \Leftrightarrow \nvdash_{\mathbf{L}} \varphi$. Thus we have proven that $\mathbf{L}$ is sound and complete to $\mathsf{A} \cup \mathsf{B}$. □

We can now define the unique class corresponding to any system. Let $\{\mathsf{C}_i\}_{i \in I}$ be the set of all classes w.r.t. which $\mathbf{L}$ is sound and complete. Let $\mathsf{C}_\infty = \bigcup_{i \in I} \mathsf{C}_i$. Now

[12]Imagine a system $\mathbf{L}$ which is sound and complete to a class consisting of one frame $\mathcal{F}$. We can double the frame to $\mathcal{F}_2$ by doubling the set T and have each relation in the original $\mathcal{F}$ occur twice, once between the original points and once between the doubles. This gives us a new frame, and for all formulas φ we have $\mathcal{F} \models \varphi$ if and only if $\mathcal{F}_2 \models \varphi$. So the system $\mathbf{L}$ is also sound and complete w.r.t. $\{\mathcal{F}, \mathcal{F}_2\}$.

C_∞ is an ordinary class of frames, since it is still a subset of the class of all frames. Secondly, C_∞ is sound and complete w.r.t. $\mathbf{L}$. Let φ be any formula.

If $\mathsf{C}_\infty \models \varphi$, then since $\mathsf{C}_i \subset \mathsf{C}_\infty$, we have $\mathsf{C}_i \models \varphi$. And since C_i is sound and complete w.r.t. $\mathbf{L}$, this implies that $\vdash_{\mathbf{L}} \varphi$.

If $\vdash_{\mathbf{L}} \varphi$, let $\mathcal{F} \in \mathsf{C}_\infty$. Then $\mathcal{F} \in \mathsf{C}_i$ for some $i \in I$. Since C_i is sound and complete w.r.t. $\mathbf{L}$, this implies that $\mathsf{C}_i \models \varphi$, and thus $\mathcal{F} \models \varphi$. Therefore, $\mathsf{C}_\infty \models \varphi$.

So C_∞ is sound and complete w.r.t. $\mathbf{L}$, and any other class C that is sound and complete w.r.t. $\mathbf{L}$ is a subset of C_∞. We define C_∞ to be *the* class that is sound and complete w.r.t. $\mathbf{L}$, and label the correspondence with $\mathsf{C}_{\mathbf{L}}$. We say that $\mathsf{C}_{\mathbf{L}}$ corresponds to the system $\mathbf{L}$.

4.6.2. *Some Examples.* With the idea of a single class corresponding to a system $\mathbf{L}$, we can look at some actual examples. But unfortunately most of them bring their own difficulties. We will dive into one of the more complex ones when proving the completeness of temporal logic. So we will omit the proofs here, and instead direct you to the book *Modal Logic* [7] for more information.

The simplest example of a correspondence is between $\mathbf{K}$ and $\mathsf{C}_{\mathbf{K}}$, which is the class of all frames, as that is what it means for $\mathbf{K}$ to be sound and complete. For many defining formulas φ, it is indeed the case that $\mathbf{K}+\varphi$ is sound and complete w.r.t. the class defined by φ. So if we look once again at $q \to \Diamond q$, which defines reflexive frames, we see that $\mathsf{C}_{\mathbf{K}+q\to\Diamond q}$ is indeed the class of reflexive frames. The system $\mathbf{K}+\Diamond\Diamond q \to q$ is sound and complete w.r.t. the class of transitive frames, and $\mathbf{K} + q \to \Box\Diamond q$ is sound and complete to the class of symmetric frames.

4.6.3. *General Lemmas.* We will end the section with two results which will be used later.

Lemma 4.26. *The class $\mathsf{C}_{\mathbf{L}+\varphi}$ is a subset of $\mathsf{C}_{\mathbf{L}}$.*

Proof. If $\mathsf{C}_{\mathbf{L}+\varphi}$, were empty, then it is a trivial subset. So let $\mathcal{F} \in \mathsf{C}_{\mathbf{L}+\varphi}$. If $\mathcal{F} \notin \mathsf{C}_{\mathbf{L}}$, then because $\mathsf{C}_{\mathbf{L}}$ is per definition maximal, $\mathbf{L}$ will not be sound and complete w.r.t. $\mathsf{C}_{\mathbf{L}} \cup \{\mathcal{F}\}$. So there exists a formula ψ such that either $\vdash_{\mathbf{L}} \psi$ and $\mathsf{C}_{\mathbf{L}} \cup \{\mathcal{F}\} \not\models \psi$ or $\mathsf{C}_{\mathbf{L}} \cup \{\mathcal{F}\} \models \psi$ and $\not\vdash_{\mathbf{L}} \psi$. If it were the latter, then from $\not\vdash_{\mathbf{L}} \psi$ we can conclude that $\mathsf{C}_{\mathbf{L}} \not\models \psi$ and thus $\mathsf{C}_{\mathbf{L}} \cup \{\mathcal{F}\} \not\models \psi$, which is contradictory. And if it were the former, then $\vdash_{\mathbf{L}} \psi$ would imply that $\mathsf{C}_{\mathbf{L}} \models \psi$. The only way that $\mathsf{C}_{\mathbf{L}} \cup \{\mathcal{F}\} \not\models \psi$ is if $\mathcal{F} \not\models \psi$. Since $\mathcal{F} \in \mathsf{C}_{\mathbf{L}+\varphi}$, this would mean that $\mathsf{C}_{\mathbf{L}+\varphi} \not\models \psi$. But since $\vdash_{\mathbf{L}} \psi$ it is also the case that $\vdash_{\mathbf{L}+\varphi} \psi$, which would mean that $\mathsf{C}_{\mathbf{L}+\varphi} \models \psi$, which is a clear contradiction.

From this we conclude that $\mathcal{F} \in \mathsf{C}_{\mathbf{L}}$, and thus that $\mathsf{C}_{\mathbf{L}+\varphi} \subseteq \mathsf{C}_{\mathbf{L}}$. □

To finish this section, we prove a general result about the relation between a system $\mathbf{K} + \varphi$ and $\mathsf{C}_{\mathbf{K}+\varphi}$, which I could not find anywhere during my research.

Lemma 4.27. *If φ defines the empty class, then the system $\mathbf{K} + \varphi$ is inconsistent, i.e. $\mathsf{C}_{\mathbf{K}+\varphi} = \emptyset$.*

Proof. Assuming that $\mathsf{C}_{\mathbf{K}+\varphi}$ is not empty, let $\mathcal{F} \in \mathsf{C}_{\mathbf{K}+\varphi}$. Then for any formula ψ,

$$\vdash_{\mathbf{K}+\varphi} \psi \Leftrightarrow \mathsf{C}_{\mathbf{K}+\varphi} \models \psi \Rightarrow \mathcal{F} \models \psi.$$

Now let us choose $\psi = \varphi$. Since $\vdash_{\mathbf{K}+\varphi} \varphi$, we have that $\mathcal{F} \models \varphi$. But that would imply that $\mathcal{F}$ is an element of the empty class defined by φ. This contradiction implies that $\mathsf{C}_{\mathbf{K}+\varphi}$ must be empty. □

The reason that this result seems almost never used is that one rarely wants to look at $\mathbf{K} + \varphi$ when φ is a formula so irregular that there are no frames in which it is necessary. But I have found a use when discussing the application of temporal logic to philosophy, see section 6.

5. Temporal Logic

In the previous section we focused on the alethic school of modal logic, and introduced a variety of topics. We can now look at the more complex temporal logic. We will mostly focus on the few differences, and explain how one can adapt the proofs of the previous section such that they apply to temporal logic.

As already mentioned in section 4.1, temporal logic deals with four distinct modals:

- *will once be in the future* (F).
- *will from now on be* (G).
- *has once been in the past* (P).
- *has been up till now* (H).

These differ from the alethic modals *is possible* and *is necessary* in their directionality. This will become clear when we formally define the modals.

5.1. Syntax of Temporal Logic.

Temporal Logic expands propositional logic with four symbols: F, G, P and H, which match their respective modal in the above list. Of these, the two symbols G and H are taken to be more fundamental. The three rules for formula construction of propositional logic has to be extended with a fourth one:

(4) If φ is a formula, then $G\varphi$ and $H\varphi$ are also formulas.

The table of abbreviations will be extended with two additional items: $F\varphi$ is an abbreviation of $\neg G \neg \varphi$ and $P\varphi$ is an abbreviation of $\neg H \neg \varphi$. Both of these symbols will have the same binding strength as $\neg$ when it comes to ordering with brackets. The symbols $\mathcal{L}_S$ and $\mathcal{L}_S^{prop}$ will have the same meaning as in modal logic.

The symbols introduced will be very symmetrical. The symbols G and H will mean the same thing but in different directions. Similar for F and G.

Definition 5.1. Each formula φ can be temporally mirrored to φ^* by replacing each instance of G with H, each instance of F with P, and *vice versa*.

It is clear that some formulas are invariant under temporal mirroring, i.e. $\varphi = \varphi^*$, as they are either pure propositional formulas, or highly symmetrical like the formula $Pq \vee q \vee Fq$.

5.2. Semantics of Temporal Logic.

In our discussion on semantics of alethic logic, we first needed to introduce the notion of a frame. Here we need to know what class of frames we take into consideration. It would be unhelpful to consider all possible frames, as some do not even come close to resembling our intuitive understanding of time. But the problem of trying to exclude these unhelpful frames is that we will inevitably also exclude some reasonable ones. For example, do we allow for circular time? While theoretically allowed by general relativity, it would invalidate the logical structure generated by the temporal models. They are not designed to describe situations where a situation can simultaneously be in the past and in the future.

In the end, we have to require two properties for the modals to make sense: they must be transitive and irreflexive, i.e. they must be a strict partial order. Since we

will be encountering frames that are strict partial orders many times, they have been given the somewhat unwieldy name **flow of time**. We will thus be talking about multiple flows of time, which can form a class of flows of time.

Because our frames now represent a timeline, they go by different names. The set T represents a time line, and consists of time points. The accessibility relation is a strict partial order, and in thus is written as $<$. When two time points $t, s \in T$ are related $t < s$, we say that t precedes s. We say that $t > s$ if $s < t$.

Models in temporal logic are exactly the same as in alethic being evaluations $\pi : T \times S \to \{0, 1\}$. These are still insufficient to be extended to an evaluation function $\Phi : T \times \mathcal{L}_S \to \{0, 1\}$, so we will again use proper models. This leads to the familiar lemma:

Lemma 5.2. *A proper model $(\mathcal{T}, \pi)$ of temporal logic defines a unique evaluation function $\Phi : T \times \mathcal{L}_S \to \{0, 1\}$, as follows: for each $t \in T$:*

$$\begin{aligned}
\Phi(t, p) &= \pi(t, p) \quad \textit{for each } p \in S \\
\Phi(t, \bot) &= 0 \\
\Phi(t, \varphi \to \psi) &= \Phi(t, \varphi) \to' \Phi(t, \psi) \\
\Phi(t, G\varphi) &= 1 \quad \textit{if for each } s \in T \textit{ with } t < s : \Phi(s, \varphi) = 1 \\
\Phi(t, G\varphi) &= 0 \quad \textit{if there exists an } s \in T \textit{ with } t < s : \Phi(s, \varphi) = 0 \\
\Phi(t, H\varphi) &= 1 \quad \textit{if for each } s \in T \textit{ with } s < t : \Phi(s, \varphi) = 1 \\
\Phi(t, H\varphi) &= 0 \quad \textit{if there exists an } s \in T \textit{ with } s < t : \Phi(s, \varphi) = 0
\end{aligned}$$

Where the symbol $\to'$ refers to the binary operator shown in table 2.

The G symbol is treated the same as the $\Box$ symbol was in alethic logic. So anything that was proven for $\Box$ in alethic logic will also apply to G in temporal logic. And if we examine how the H symbol is treated, is actually also equivalent to the way the $\Box$ symbol was used. The choice between the left or right side of an operation is arbitrary. This will generally be how one adapts proofs from alethic logic to temporal logic.

Proof. We can adapt the proof of lemma 4.3 by first replacing $\Box\varphi$ with $G\varphi$ and R with $<$, from which we can conclude that the two extensions Φ and $\hat{\Phi}$ must evaluate $G\varphi$ to the same value. Then by repeating the argument with $H\varphi$ and replacing R with $>$, we see that the two extension must evaluate $H\varphi$ to the same value. Again, the total number of occurances of G and H is finite, so repeating the argument once again gives $\Phi = \hat{\Phi}$. □

In modal logic this lemma was followed by a section on modal tautologies. All of these can be shown to be the case for temporal logic by following the same steps. First, we replace $\Box$ with G and R with $<$. The argument should then still hold. Secondly, we replace $\Box$ with H and R with $>$. This should likewise hold. Then we can conclude that the property also works for temporal logic.

We end with a quick mention that the symbols $\Box$ and $\Diamond$ are also used in temporal logic. They cannot have the same meaning as in alethic logic, but the idea of necessary and possible is also applicable in the context of time. When at a time point t one says that something is possible, it is more precise to say that there exists an accessible time point, i.e. a point s such that $s < t$, $t = s$ or $t < s$, where the thing is true. This is captured by $\Diamond\varphi$, which is defined as $P\varphi \vee \varphi \vee F\varphi$. The concept of φ being

necessary then becomes $\Box\varphi$, which is an abbreviation for $H\varphi \wedge \varphi \wedge G\varphi$. It is clear that

$$(\Diamond\varphi)^* = \Diamond\varphi^* \text{ and } (\Box\varphi)^* = \Box\varphi^*.$$

5.3. **Deduction in Temporal Logic.** When we turn to the question on what deduction system is suited for temporal logic, we can first try to use the theory of section 4.6, since we are seeking to axiomatize a class of frames. But unfortunately, we run into the problem of undefinability. It was already mentioned that in alethic logic there is no formula which defines the irreflexivity property. And that argument can quite easily be generalised, implying that in no modal logic the class of irreflexive frames can be defined.

So while the theory of axiomatization set out in section 4.6 fails here, that does not mean we are doomed. A deduction system for temporal logic does exist, but it is not one prescribed by axiomatization: it is wholly unique. That means that the soundness and completeness proofs have to be restated to take the new details in considerations. But first, let us introduce the base deduction system of temporal logic **B**.

Definition 5.3. B is a system that expands the A. Church's system with two additional deduction rules and four additional axiom schemes.
Additional deduction rule:

G-Generalisation: As input we take one formula of the form α and we output $G\alpha$.

H-Generalisation: As input we take one formula of the form α and we output $H\alpha$.

Axiom scheme:

(4) $G(\alpha \to \beta) \to (G\alpha \to G\beta)$.
(5) $H(\alpha \to \beta) \to (H\alpha \to H\beta)$.
(6) $p \to GPp \wedge HFp$.
(7) $Gq \to GGq$.

Axiom scheme 4 and 5, also known as the distribution axioms, regulates the G and H symbols respectively. Axiom scheme 6, the converse axiom, defines how G and H interact with each other. Axiom scheme 7 is the defining formula for transitivity. From the examples in section 4.6 we can see that this will limit the frames to only the transitive ones.

We again have the notation of adding an axiom φ to a system **L**, which is still denoted by $\mathbf{L} + \varphi$. And by restating that the deduction theorem still holds with the modal definition of $\Sigma \vdash_{\mathbf{B}} \varphi$, we can again use the propositional results about consistent and inconsistent theories.

5.4. **Soundness and Completeness of Temporal Logic.** We will now turn to the sound and completeness of temporal logic. Since we are dealing with a class of frames that is undefinable, we cannot use any results of axiomatization. So we are required to repeat the proofs from section 4.5. But while until now we could reuse most of these with only minor alterations, we will come across a difficult problem if we attempt to do so for the completeness proof. We will need to introduce the tool of bounded morphisms to solve this problem.

Theorem 5.4. *Temporal logic is sound and complete.*

Each of the two halves will be proven separately.

5.4.1. *Soundness of Temporal Logic.* The soundness proof raises no difficulties in its translation from modal logic.

Lemma 5.5. *Temporal Logic is sound, i.e.* $\vdash \varphi \Rightarrow \models \varphi$

Proof. The same argument for lemma 4.19 holds here, where we replace $\Box$ once with G and once with H. We only have to mention that $\models \varphi$ implies both $\models G\varphi$ and $\models H\varphi$. $\Box$

5.4.2. *Completeness of Temporal Logic.* The completeness proof will be more difficult, as simply replacing some symbols will not suffice.

Lemma 5.6. *Temporal Logic is complete, i.e.* $\models \varphi \Rightarrow \vdash \varphi$.

The idea of the proof is still the same. We will show the contrapositive by using the Canonical Model. But first it must be updated to the new terminology. Let $\mathcal{S}$ be the set of maximally consistent theories. On this set we define the relation R_c as follows:

$$\Sigma R_c \Delta \text{ if and only if } \forall_{\alpha \in \Delta} F\alpha \in \Sigma.$$

We can then check that the three properties of maximally consistent sets are also true for temporal logic:

(1) $\Sigma \vdash \varphi$ if and only if $\varphi \in \Sigma$.
(2) $\varphi \notin \Sigma$ if and only if $\neg\varphi \in \Sigma$.
(3) $F\varphi \in \Sigma$ if and only if $\exists_{\Delta \in \mathcal{S}} \Sigma R_c \Delta$ and $\varphi \in \Delta$.

Of these, (3) will become important, and therefore it will be called the **future existence property**. It was a key component in the "Truth coincides with membership" property in alethic logic, but in temporal logic, we need something similar, now with regard to the **past existence property** and $P\varphi$.

Lemma 5.7. *For any* $\Sigma \in \mathcal{S}$*, and any formula* φ*:*

$$P\varphi \in \Sigma \textit{ if and only if } \exists_{\Delta \in \mathcal{S}} \Delta R_c \Sigma \textit{ and } \varphi \in \Delta.$$

Proof. We claim that $\Delta R_c \Sigma$ if and only if $\forall_{\alpha \in \Delta} P\alpha \in \Sigma$.

First assume that $\Delta R_c \Sigma$. If it is not the case that $\forall_{\alpha \in \Delta} P\alpha \in \Sigma$, then there exists an $\alpha \in \Delta$ such that $P\alpha \notin \Sigma$. Since Σ is a maximally consistent theory, we find that $\neg P\alpha \in \Sigma$. By definition, $\Delta R_c \Sigma$ means that $F\neg P\alpha = \neg GP\alpha \in \Delta$. But because of the converse axiom and $\alpha \in \Delta$, we have $\Delta \vdash GP\alpha$. This contradiction means that $\forall_{\alpha \in \Delta} P\alpha \in \Sigma$.

Assume that $\forall_{\alpha \in \Delta} P\alpha \in \Sigma$, but not $\Delta R_c \Sigma$. Then there exists an $\alpha \in \Sigma$ such that $F\alpha \notin \Delta$. Using the same argument as above, we can conclude that $\neg HF\alpha$ and $HF\alpha \in \Sigma$. This contradiction means that $\Delta R_c \Sigma$.

With this extra property of the accessibility relation, we can use the same argument as in 4.22, but replacing $\Box$ and $\Diamond$ with H and P, and switching the order of $\Sigma R_c \Delta$ to $\Delta R_c \Sigma$. $\Box$

Now that we have both existence properties, we can use the same proof to show that the canonical model also has the "Truth coincides with membership" property within temporal logic. While in alethic logic we were basically done, in temporal logic we run into a fundamental problem. The canonical model is not a flow of time, as it will have reflexive points, i.e. $\Sigma \in \mathcal{S}$ with $\Sigma R_c \Sigma$. That is why we still use the general relation symbol R_c instead of the more specific $<$. However, the canonical model can be transformed into a flow of time. But first we need to establish that canonical model is transitive.

Lemma 5.8. *The frame $(\mathcal{S}, R_c)$ of the canonical model in temporal logic is transitive.*

Proof. Let $\Sigma R_c \Delta$ and $\Delta R_c \Lambda$ but not $\Sigma R_c \Lambda$. Then there must be a formula $\alpha \in \Lambda$ such that $F\alpha \notin \Sigma$ and thus $\neg F\alpha = G\neg\alpha \in \Sigma$. Because of the axiom $G\varphi \to GG\varphi$, $G\neg\alpha \vdash_{\mathbf{B}} GG\neg\alpha$, and thus $GG\neg\alpha = \neg FF\alpha \in \Sigma$. But since $\Delta R_c \Lambda$ and $\alpha \in \Lambda$, $F\alpha \in \Delta$, and because $\Sigma R_c \Delta$, this implies that $FF\alpha \in \Sigma$. This contradiction implies that $\Sigma R_c \Lambda$, and thus $(\mathcal{S}, R_c)$ is transitive. □

We will solve the issue by altering the canonical model such that it no longer has any loops, but is still transitive and has the important "Truth coincides with membership" property. To show the latter, we are going to use a powerful tool in modal logic, the bounded morphisms.

Bounded morphisms are the structure preserving mappings of proper models. While these mappings are very interesting and have many uses, delving into the theory would make the thesis even longer. Instead, I will refer to P. Blackburn, M. De Rijke and Y. Venema, *Modal Logic* [7]. According to their definition 2.10 a bounded morphism between proper models is the following:

Definition 5.9. A bounded morphism is an map between two proper models $(T, <, \pi)$ and $(T', <', \pi')$, where $t \in T \mapsto t' \in T'$, satisfying:

- $\pi(t, q) = \pi'(t', q)$ for all $q \in S$ (Model preserving).
- $t < s$ implies that $t' <' s'$ (Forwards preserving).
- $t' <' x$, where $x \in T'$ implies that there is an $s \in T$ with $t < s$ and $s' = x$ (Backwards consistency).

For a morphism, something like the model preserving and forwards preserving properties are expected, since we want the proper models to be preserved when applying the morphism. But the backwards consistency property is less intuitive. It is required to ensure that when a proper model is mapped onto another proper model, the limiting properties of the original accessibility relation are also preserved.

Using this definition, we get from proposition 2.14 in [7] the following lemma:

Lemma 5.10. *If there exists a bounded morphism between two proper models $(T, <, \pi)$ and $(T', <', \pi',)$, then for any formula φ, $\Phi(t, \varphi) = \Phi'(t', \varphi)$, i.e. the truth value of formulas is preserved.*

The proof goes through induction on the length of the formula φ, but is left out for brevity (see [7]).

We now define a flow of time such that there is a bounded morphism between it and $(\mathcal{S}, R_c, \pi_c)$. This will be the canonical flow of time $(\mathbb{Z} \times \mathcal{S}, <, \pi_c)$. The elements of $\mathbb{Z} \times \mathcal{S}$ is noted as $\Sigma_n := (n, \Sigma)$. We define every Σ_n to be a copy of the original Σ, so they contain the same elements. This way, we can use the same definition of π_c as for the canonical model, namely $\pi_c(\Sigma_n, p) = \chi_{\Sigma_n}(p) = \chi_\Sigma(p)$.

The relation $<$ is the product relation of R_c on $\mathcal{S}$ and $<$ or $\mathbb{Z}$, i.e. $\Sigma_n < \Delta_m$ if and only if $\Sigma R_c \Delta$ and $n < m$. Since both R_c and $<$ on $\mathbb{Z}$ are transitive, $<$ will also be transitive.

An obvious candidate for a bounded morphism from $(\mathbb{Z} \times \mathcal{S}, <, \pi_c)$ to $(\mathcal{S}, R_c, \pi)$ would be the projection $\mathbb{Z} \times \mathcal{S} \to \mathcal{S}$.

Model preserving: Since Σ_n and Σ contain the same elements, $\pi_c(\Sigma_n, p) = \chi_{\Sigma_n}(p) = \chi_\Sigma(p) = \pi_c(\Sigma, p)$.

Forwards preserving: From the definition of $\Sigma_n < \Delta_m$, we can directly conclude that $\Sigma R_c \Delta$.

Backwards consistency: Let us look at Σ_n, which is projected to Σ. If $\Sigma R_c \Delta$, then Δ_{n+1} has both $\Sigma_n < \Delta_{n+1}$ and $\Delta_{n+1} \mapsto \Delta$. So the mapping has the backwards consistency property.

With this, we can proof the "Truth coincides with membership" property of the canonical flow of time.

Lemma 5.11. *The canonical flow of time has*

$$\Phi(\Sigma_n, \varphi) = 1 \text{ if and only if } \varphi \in \Sigma_n.$$

Proof. Let Φ be the extension from the canonical flow of time and Φ_c from the canoncial model. Since there exist a bounded morphism between the two, we can use lemma 5.10 to show that

$$\Phi(\Sigma_n, \varphi) = \Phi_c(\Sigma, \varphi) = \chi_\Sigma(\varphi) = \chi_{\Sigma_n}(\varphi),$$

which is what needed to be shown but rephrased. □

Now we can finally prove lemma 5.6.

Proof of Lemma 5.6. We will prove the contrapositive: if $\not\vdash \varphi$, then $\{\neg\varphi\}$ is consistent. We extend the set $\{\neg\varphi\}$ to the maximally consistent set Σ. Then in the proper flow of time, using lemma 5.11, we find that $\Phi(\Sigma, \neg\varphi) = 1$. From this we can conclude that $\Phi(\Sigma, \varphi) = 0$. So $\not\models \varphi$. □

5.5. **Axiomatization.** We end this section about temporal logic by going back to section 4.6. If we look at the proofs, we quickly see that these do not rely on the specifics of alethic logic, except for the examples. In particular, both lemma 4.26 and lemma 4.27 are also true when **K** is replaced with **B**. Furthermore it is still the case that most extended systems need to be checked individually. For example, one can look at the axiomatization of linear flows of time[13], which are flows of time that are total orders.

Let $\varphi = PFq \to \Diamond q$. One can prove that φ defines the class of flows of time that are non-branching to the future, i.e. for every time point t, any two distinct points s, r in t's future have either $s < r$ or $r < s$. The temporal mirror φ^* defines non-branching to the past. By taking the conjunction of the two we find the defining formula for linear flows of time: $\varphi_{lin} = \varphi \wedge \varphi^*$.

With this, we have introduced everything that we need to know to prove the two main results of this thesis. While some more concepts of modal logic will be mentioned in the upcoming sections, they will not be proven so exhaustively as we have done up to this point.

6. Applying to Philosophy

With the theory of temporal logic now firmly established, we can focus ourselves on the second objective of this thesis, namely applying this theory to the problem of time. As mentioned in the Introduction, the application will be done in two parts. Here we will be looking at what philosophers call the problem of time, where as in the next chapter we will look at what physicists call the problem of time.

But before we can start, we need to discuss some caveats. While this is a mathematical thesis, this section will be quite different from the rest. Instead of giving strict mathematical proofs, we will look into a different discipline, written to be understood

[13]Technically we are dealing with component wise linear flows of time, but that distinction has no practical ramifications.

by my fellow mathematicians. So before we can delve into applying temporal logic to any aspect of philosophy, we need some background as to how formal logic is used in philosophy.

Unfortunately, as a bachelor student of mathematics, I have no experience in the field of philosophy. I have read a multitude of textbooks which introduced me to at least the philosophy of space and time: B. Daiton *Time and Space* [8], C. Callender *What Makes Time Special?* [9] and B.L. Curtis and J. Robson, *A Critical Introduction to the Metaphysics of Time*. But I knew that these books would not give me all the answers I needed. So I contacted two experts in the field of philosophy: Natalja Deng[14] and Neil Dewar[15]. While they are my sources, anything wrong here will be because of mistakes I myselve made while writing this thesis. I also had a fruitful conversation with Jeremy Butterfield [16], which helped me orientate on the subject and provided helpful resources.

6.1. **What is the problem of time.** The specific field of philosophy that deals with questions about time is, unsurprisingly, called the philosophy of time. The different points they argue are collectively called the problem of time. But as a simple overview, the problem of time can be reduced to two categories: *What is time?* and *Why is time the way it is?*[17]. These broad questions encompass many different aspects of time that philosophers seek answers to. To mention a few,

What is time:

- Is time universal, i.e. do I experience time the same way everyone else does?
- Does the present exist, or is it an illusion?
- Does time itself flow or does only the present flow, or neither?

Why is time the way it is:

- Why is time one directional?
- Why is the past so distinct from the future?
- If time is ultimately an illusion, why do we experience time?

This is clearly a broad subject, so instead of showing how applying temporal logic would interact with each one of these question, I will dive a bit more deeply into one question: does time exist? And I will focus only on a single, but very important paper discussing this question: J. M. E. McTaggart, *The Unreality of Time*. Every book on the subject introduces McTaggart as the father of the modern philosophy of time. He introduced a pivotal distinction in the way we speak about time, which has shaped the discussion ever since it was published in 1908. It has spawned two schools of thought, which still argue for their view up until today. With the amount of attention it has already been given, it is a well understood argument, which eases the daunting task of applying a logical formalism.

6.2. **Philosopher's usage of formal logic.** Before we delve into McTaggart, we need to know how philosophers use logic. But I quickly learned that the question

[14]Underwood International College, Yonsei University, Seoul/Incheon, South Korea. Personal Homepage: `http://nataljadeng.weebly.com/`

[15]Munich Center for Mathematical Philosophy, LMU, Münich, Germany. Personal Homepage: `https://neildewar.wordpress.com/`

[16]Trinity College, Cambridge, UK. Email: jb56@hermes.cam.ac.uk

[17]This distinction is a bit arbitrary and of my own making. One can easily rephrase the questions to make it seem like they belong to the other category. So do not put too much stock in this distinction.

of how philosophers use logic is far too broad to be answered in any single paper. So instead I will give you only a small summary. In general, there are two distinct notions the question covers. First is the philosophy of logic, and second is using a logical formalism when giving a philosophical argument. The former deals with questions such as what makes a formula true, what logic underlies our universe, how can one deal with logical paradoxes, and many more. The latter seeps into every aspect of philosophy, and hence to answer the question one would need to know the inner workings of the field of philosophy. This is where most of my time with the experts was spent.

6.2.1. *Philosophy of Logic.* As an example of the philosophy of logic, let us examine the famous liar paradox: A person says: "I am lying." Is he speaking the truth, or is he lying? If he is lying, then he said the truth, if he is telling the truth, then he is lying. The question then becomes: what does this paradox mean for logic? The Stanford Encyclopedia of Philosophy (SEP) [10] has, among other things, this to say:

> From time to time, the Liar [paradox] has been argued to show us something far-reaching about philosophy. For instance, Grim has argued that it shows the world to be essentially 'incomplete' in some sense, and that there can be no omniscient being. McGee and others suggest that the Liar [paradox] shows the notion of truth to be a vague notion. Glanzberg holds that the Liar [paradox] shows us something important about the nature of context dependence in language, while Eklund holds that it shows us something important about the nature of semantic competence and the languages we speak. Gupta and Belnap claim that it reveals important properties of the general notion of definition. And there are other lessons, and variations on such lessons, that have been drawn.
>
> (Internal citations omitted)

This, while understandingly being a bit shallow in its explanation[18], is an example of philosophy of logic. They put forth arguments for what the nature of logic is, what it can and cannot do and what effect this has on something as concrete as human language.

6.2.2. *Logic in Philosophy.* While the study of philosophy of logic is very interesting, it is not where the rest of this thesis will focus on. Instead we will look at the second way logic can be used in philosophy: applying logical formalism to philosophical arguments. You have already seen a simple example of what a philosophical argument is in the description of the liar paradox. It is a well formulated and understandable argument for why a certain idea is true. Unlike a mathematical proof, is it usually written in a natural language. This makes it less precise, but more flexible to deal with the more diverse issues philosophy is working with.

But the lack of precision does come at a cost, as I found out myself. We will soon talk about McTaggart's argument for the unreality of time, and the argument has been subject to one discussion for the past hundred years: What do the words he wrote down actually mean? To combat this issue, some authors include a logical formalism when setting out their arguments. As one can see by scanning through the SEP paper on the liar paradox [10], many different formalisms are used, but none in

[18]This section was chosen for its clarity for people with a limited experiance in the field of philosophy. But that was then necessarily not the most thorough explanation.

the same way that a mathematician would use it. To understand this, I asked Dewar specifically, as he has attempted to apply a strict mathematical formalism before. He mentioned four issues one encounters when using a strict formalism in philosophy:

(1) **Very little insight is gained from using a formalism.** What lays in the heart of philosophy is broadening one's understanding of how things can work. And strictly adhering to a formalism does not further that goal.
(2) **There would be much fruitless debate on which formalism to use.** Similar to mathematical debates between logicism and intuitionism etc., philosophers would have to spend a significant time discussing which formalism is the best fit. Since the current system of using natural language works fine, this would be time spent not working on furthering understanding.
(3) **You would have to decide in advance which formalism to use, which is sometimes impossible to determine.** Fitting an existing argument into a formalism that it was not built for, will most likely be difficult. This almost requires that before any work furthering understanding can be done, a decision must be made on which formalism to use. And at that point, there is little information to understand which formalism will turn out to be the right choice.
(4) **Natural language is not yet comparable to any existing formalism.** The arguments set out in all of philosophy will use almost every feature of natural language. And research into natural language has yet to result into one formalism that perfectly describes all its features. These arguments would be unavailable to philosophy if we restricted ourselves to only formalisms.

It is also good to mention that philosophy, similar to physics, mostly deals with our world. While a logical formalism can be shown to be internally consistent, we have not yet proven any one to be the correct one for describing the logical structure of our universe. This might even be impossible, as the entirety of the universe could be too large to be observed. So even claiming that propositional logic works for our universe is a postulate. So demanding philosophy to be restrained by propositional logic would limit one's ability to do philosophy of logic.

6.2.3. *How to Apply Temporal Logic.* With all these points against formalisms, it might seem foolish to set forth applying temporal logic to philosophy. But I think there is a bit more to say. While it is clearly unreasonable to expect every single philosophical argument to fit neatly into a formalism, one can still look at what insight can be gathered from specifically applying a formalism to a specific argument, such as applying temporal logic to McTaggart's argument for the unreality of time. But before we can do exactly that, we need to understand McTaggart's argument in words.

6.3. **McTaggart's argument.** McTaggart was a man of contradictions. Very much rooted in the Hegelian ideas of the time, he was simultaneously critical on much of Hegel's work. While most of the time politically conservative, he was an advocate of woman suffrage and while atheistic, he believed in the immortality of the human soul. In our conversation, Butterfield described him as an idealist, believing among his more reasonable ideas that the universe was so complex that it cannot be the case that our everyday understanding of it was correct.

Within this framework, he set out to show that time as we experience cannot be real: *The Unreality of Time* [11]. And he came with an argument that has

been difficult to refute for the past century. But on top of that, he introduced new concepts which are still important to this day: the A and B series. In this section I will explain McTaggart's argument. While the original paper goes much further in depth, considering more possible objections, the core of the argument will hopefully come across. An overview of his argument is as follows:

- There are two ways we can look at time, the A series and the B series.
- Of the two, the A series is more fundamental and it is necessary for time to exist.
- The A series is absurd, as it results in an infinite regression.
- Thus, time cannot be real.

6.3.1. *The A and B series.* When we look at the linguistic treatment of time, we can distinguish two different patterns of speech. In the case of days, you can speak of yesterday, today and tomorrow, and you can speak of 31$^{\text{st}}$ of May, 1$^{\text{st}}$ of June and 2$^{\text{nd}}$ of June, 2020. While subtle, there is a distinct difference between the relative nature of the first and the absolute nature of the second. At whatever moment I pronounce 1$^{\text{st}}$ of June, 2020, it will always refer to the same date. But when one encounters "today" without any context, we are left in the dark as to what is meant.

To formalise this difference, McTaggart introduced the notion of the A and B series. These are orderings on the events of the universe. What these events are is a contested issue[19], but for the moment let them be the points in a spacetime. The A series is the ordering which runs from the far past to the near past to the present to the near future to the far future. Meanwhile, the B series runs from what is earlier to what is later.

6.3.2. *The necessity of the A series.* McTaggart claims that the A series is necessary for time to be real, in that between the different ways to describe time, the A series is the fundamental way to look at it. He argues as follows. It is clear that time requires change. While this premise has some hidden assumptions, for this thesis we will accept it. Now the question is, how can change manifest itself? According to McTaggart, there is no change in the B series, since the order of events is static. Whether we are looking at it from the birth of Julius Caesar, from the start of World War I, from the first day of the new millennium or a moment in the year 2222, two events in the B series will always have the same relation to each other. Therefore, while a useful construct, the B series does not contain the origin of change, and therefore is not fundamental to time.

The A series, on the other hand, does allow for change. An event can at some point be in the future, at another in the present and yet at a later point in the past. To show that the A series is fundamental, let us examine a series with fewer properties, something more fundamental. The C series is similar to the A series, but it lacks directionality. Immediately, we have the problem that change no longer occurs. The fundamental ordering on which the A series is built, namely that event y comes between x and z, will be the same regardless of at which moment we are

[19]Let me briefly mention three views on this topic. McTaggart says events are "the contents of a position in time." Dainton [8] defines events as "a happening or process, typically extended over both time and space." Callender [9] says "We represent the set of all events with a four-dimensional smooth and connected manifold M.", and then later mentions that M is our spacetime manifold. These three view are fundamentally incompatible, and this lays at the root of some counterarguments.

looking at them. So, the A series is sufficient for time, and the C series is insufficient for time. Therefore, the A series must be necessary.

6.3.3. *The absurdity of the A series.* Now let us inspect a single event in the A series. McTaggart uses the death of Queen Anne. An event can't be both past, present and future. The very nature of being one excludes the two others. Yet, when we look at the date August 1st, 1714, the death would be in the present, when we look at the date January 1st, 1714, it would be in the future, and when we look at the date December 1st, 1714, it would be in the past.

Now, a quick response you could make is to say: of course it doesn't have the three properties *simultaneously*, but in *succession*. On August 1st, 1714, the event has been in the future, is in the present and will be in the past. On January 1st, 1714, it is in the future, will be in the present and will be in the past. And on December 1st, 1714, it is in the past, has been in the present and has been in the future.

But, according to McTaggart, you would have made a vital mistake in this response. For you say that one "being future" is in the past, and another "being future" is in the future. Therefore, we need to have an ordering which allows us to label an A series as being in the past, in the present or in the future. Let this ordering be called the A' series. Now the problem arises that this A' series has the same issues. Each series has all of pastness, presentness and futurity. We could argue that those series have the properties in succession, but for that to be the case, we need to order the A' series in a A" series that will have the same issue.

With this, McTaggart argues, it is clear that the reality of the A series is absurd, because it requires events to posses the attributes of being past, being present and being future. So the A series must be rejected.

6.3.4. *The Unreality of Time.* McTaggart has shown that the A series is both necessary for time, and yet absurd. The only way this can be resolved, is to reject the very notion of time. Time, therefore, has to be an illusion, in some way constructed by our consciousness.

While this argument may be convincing at first, there is a lot of discussion on almost every point of the argument. As Deng explained in our conversations, there are generally speaking two camps on how one can refute McTaggart's argument. The first, called the A theorist, will have issue with the absurdity of the A series. There are many more ways to explain how past, present and future can be in succession without requiring an infinite regression. The second camp, the B theorist, have similar issues with the absurdity of the A series as the A theorist, but also have grievances with the argument for the necessity of the A series. McTaggart claims that change can't occur within the B series, but has a very strict definition of what change is. For example, the death of Queen Anne, which unequivocally changes something about Queen Anne, would not be a change in McTaggart's explanation of change. The B theorist will claim that the B series is fundamental to time.

6.4. **Analysing McTaggart's Argument.** Now that we know both how philosophers use logic, and have gone in depth what McTaggart's argument is, we can see how it can be analysed through formal logic. The analysis will consists of two parts, my attempt to formalise the argument with minimal deviation, and seeking what can be said about McTaggart's argument from the point of view of temporal logic.

6.4.1. *Formalising.* Unfortunately for this paper, I must admit that my attempts to formalise McTaggart's argument as a formal proof within temporal logic such as in section 5.3 did not result into anything useful. It was only later, when I discussed the usage of logic with the experts that I learned that I came across the same problems as those mentioned in section 6.2.2. The argument, as written in the original paper, was not intended to be formalised in the way I was attempting.

At one point I thought about trying to prove the two sub-parts, *the necessity of the A series* and *the absurdity of the A series*, from scratch, using McTaggart as a guide. But this would not guarantee that I stay true to McTaggart's argument. If I then came to the conclusion that the argument does not hold up, this conclusion may well follow from my wrong translation into temporal logic.

So, instead of forcing myself to look at temporal logic for this part, I loosened to a form of propositional logic. This has not only resulted in a schematic dissection (figure 3, explanation of symbols in table 3), but was an important tool in my understanding of McTaggart's argument. While it could have been written as a formal proof within propositional logic, I opted for a more lenient approach. This resulted in something far more readable. It is important to note that the basis of this dissection was the original text, rather then the summary given in section 6.3. Therefore it refers to concepts and ideas not commented upon in the summary.

The overview consist of 39 numbered lines, each consisting of two parts: the formula and a descriptor at the end. There are four descriptors:

- $TAUT$. The formula is a propositional tautology. It only appears once, as McTaggart did not refer to many tautologies, except for in the beginning, when he pondered the possibility that the existence of the A series were a mere illusion.
- $CLAIM$. The formula is an assumption or indivisible. I call something indivisible as a compromise of formalising the philosophical argument. While for some of these formulas the argument went into more detail, it was difficult to capture that into the dissection. So then a reasonable endpoint was chosen and designated as an assumption. It is important to note that labelling something as an assumption is not a judgement call, as most are are quite defensible.
- $LEMMA$. The formula is stated here, and proven at a later point. This ordering of first mentioning the conclusion and then showing the proof is how the original argument was ordered.
- $PROOF$. The descriptions of which lines can be combined using propositional deduction rules to conclude the corresponding $LEMMA$.

For brevity, I will not walk through the entirety of the dissection, explaining every single line. But to illustrate how it is meant to be read, I will go through lines 16-21.

16. We are going to prove that "time exists" implies "the existence of an A series that fundamentally represents our time." This is the second half of the necessity of the A series part of the argument: explaining why the A series is minimal.
17. We claim that time can only exist if there is change.[20]
18. We claim that change must manifest itself in the A series, the C series or the C series extended with direction.

[20]This claim appears in an earlier line, but is duplicated to have each sub-proof be self-contained.

1.	$T \leftrightarrow A, \neg A \vdash \neg T$	$TAUT$
2.	$T \rightarrow A, \neg\text{Real}(A) \vdash \neg T$	$LEMMA$
3.	$\neg\text{Real}(A) \rightarrow \neg A$	$CLAIM$
4.	$(3) + (1) \Rightarrow (2)$	$PROOF$
5.	$T \leftrightarrow A$	$LEMMA$
6.	$A \rightarrow T$	$LEMMA$
7.	$(A \vee B) \rightarrow T$	$CLAIM$
8.	$T \rightarrow \neg B$	$LEMMA$
9.	$T \rightarrow \Delta$	$CLAIM$
10.	$B \rightarrow \neg\Delta$	$LEMMA$
11.	$\Delta \wedge B \rightarrow \Delta_{B1} \vee \Delta_{B2} \vee \Delta_{B3} \vee \Delta_{B4}$	$CLAIM$
12.	$B \rightarrow \neg(\Delta_{B1} \vee \Delta_{B2} \vee \Delta_{B3} \vee \Delta_{B4})$	$CLAIM$
13.	$(11) + (12) \Rightarrow (10)$	$PROOF$
14.	$(9) + (10) \Rightarrow (8)$	$PROOF$
15.	$(7) + (8) \Rightarrow (6)$	$PROOF$
16.	$T \rightarrow A$	$LEMMA$
17.	$T \rightarrow \Delta$	$CLAIM$
18.	$\Delta \rightarrow A \vee C \vee C_+$	$CLAIM$
19.	$C \rightarrow \neg\Delta$	$CLAIM$
20.	$C_+ \leftrightarrow A$	$CLAIM$
21.	$(17) + (18) + (19) + (20) \Rightarrow (16)$	$PROOF$
22.	$(6) + (16) \Rightarrow (5)$	$PROOF$
23.	$\neg A$	$LEMMA$
24.	$A \rightarrow A_{rel} \vee A_{qual}$	$CLAIM$
25.	$\neg A_{rel}$	$LEMMA$
26.	$A_{rel} \rightarrow (\epsilon \rightarrow P_\epsilon \vee N_\epsilon \vee F_\epsilon)$	$CLAIM$
27.	$\neg(\epsilon \rightarrow P_\epsilon \vee N_\epsilon \vee F_\epsilon)$	$LEMMA$
28.	$\neg(P_\epsilon \wedge N_\epsilon \wedge F_\epsilon)$	$CLAIM$
29.	ϵ	$CLAIM$
30.	$P_\epsilon \vee N_\epsilon \vee F_\epsilon \rightarrow P_\epsilon \wedge N_\epsilon \wedge F_\epsilon$	$LEMMA$
31.	$\neg(P_\epsilon \vee N_\epsilon \vee F_\epsilon \rightarrow P_\epsilon \wedge N_\epsilon \wedge F_\epsilon) \rightarrow \infty$	$CLAIM$
32.	$\neg\infty$	$CLAIM$
33.	$(31) + (32) \Rightarrow (30)$	$PROOF$
34.	$(28) + (29) + (30) \Rightarrow (27)$	$PROOF$
35.	$(26) + (27) \Rightarrow (25)$	$PROOF$
36.	$\neg A_{qual}$	$LEMMA$
37.	$(25, 26, \ldots, 35)[A_{qual}/A_{rel}] \Rightarrow (36)$	$PROOF$
38.	$(24) + (32) + (36) \Rightarrow (23)$	$PROOF$
39.	$(1) + (5) + (23) \Rightarrow \neg T$	$PROOF$

FIGURE 1. Dissection of McTaggart's argument for the unreality of time. Usage of symbols are explained in table 3.

A, B	The existence of an A/B series that fundamentally represents our time.
T	Time exists
Real(A)	The A series is not an illusion.
Δ	Change occurs.
$\Delta_{B1}, \ldots, \Delta_{B4}$	Four different aspects in which change can manifest in the B series.
$A_{\text{Rel}}, A_{\text{Qual}}$	The Past, Present, Future distinction brought by the A series is to be thought of as relations or qualities respectively.
C	The existence of a C series that represents our time.
C_+	The C series extended with direction.
ϵ	There exists an event in time.
$P_\epsilon, N_\epsilon, F_\epsilon$	The event ϵ can be described as Past, Present or Future.
∞	There is an infinite regression.

TABLE 3. Explanation of the symbols.

19. But, we claim, in the C series there is no change.
20. And, we claim, the A series is equal to the C series extended with direction.
21. From this, we can conclude that the change must manifest in the A series, as it cannot manifest from the C series, and the statement is equivalent to saying that the C series when extended with direction manifest change. Then repeatedly using substitution, the original $LEMMA$ is gotten.

As mentioned in the explanation of the $CLAIM$ descriptor, there are two types of claims. The first are those whose logical basis did not fit neatly into the formalism. An example is line 3, where the explanation of the difference between being a mere illusion and actually existing is explored in some details. But that is an argument on definitions, which is difficult to portray in propositional logic. The second type of claims are true claims in McTaggart argument. Sometimes they go unstated, such as in line 11. The four different ways change could manifest were listed, and it was implied that this exhausted all the options.

The result of phrasing McTaggart's argument in this half formalism is to show that the argument does not rest upon fallacies. The underlying logic is sound, since it was expressible in propositional logic. So for the validity of McTaggart's argument, one needs to look at the assumptions he made. But unfortunatly, this still had not much to do with temporal logic, so I leave it as a question for further research. The tool has located the claims McTaggart is based upon, and it was of great help in my understanding of the finer complexities of McTaggart's argument.

6.4.2. *Strict Temporal Logic.* As the previous section mentioned, my first attempt to apply temporal logic to McTaggart's argument ended with no use of temporal logic. To aid the research, I looked into the way different authors had already written on McTaggart's argument and strict temporal logic. And while many philosophers do seem to use temporal logic, it is usually only its trappings, i.e. the syntax is used to describe statements but none of the theory is applied[21]. And while it is helpful in combating the ambiguity problem I have already noted, it is not what I was looking for.

In the end, I found one paper: H. Rostomyan, *McTaggart's Argument for the Unreality of Time: A Temporal Logical Analysis* [13]. And how Rostomyan did the

[21]For an example, look at E.J. Lowe *McTaggart's Paradox Revisited* [12].

analysis is very interesting for my thesis. But first we need to discuss what his analysis entailed. He covers the two halves of McTaggart's argument separately.

For the necessity of the A series, he comes to the conclusion that it holds up, although with some caveats. Unfortunately for this thesis, his argument relies on a different logic that does not seem to have a name, but I would call precedent logic. Because of this, I will only sketch the argument here. The only thing that needs to be noted about precedent logic is that its proper models are also proper models for temporal logic.

Change, as he defines it, has two levels: in the proper models and when the proper models are expressed by the semantics of the logic. A changing proper model has at least one atomic proposition that at one point is true and at another point is false. Such a model then expresses change within the semantics of a logic if there is a formula within the logic that is true at one point, and false at another. He claims that the A series corresponds to temporal logic, from which he proves that changing proper models express temporal logic change. The B series corresponds to precedent logic, from which he proves that changing proper models do not express precedent logical change.

The second part, the absurdity of the A series, keeps to temporal logic. He restates McTaggart's notion that an event has all of presentness, pastness and futurity, which from now on will be called McTaggart's conclusion, to the following:

$$\text{For every formula } \varphi, \models \varphi \to P\varphi \wedge F\varphi. \tag{6.1}$$

He then proves the following proposition:

Proposition 6.1. *In temporal logic, it is not the case that for each φ,*

$$\models \varphi \to P\varphi \wedge F\varphi.$$

Proof. Let us look at the flow of time $(\mathbb{Z}, <)$, with the ordinary strict ordering $<$, together with the evaluation function π, such that $\pi(t, q) = \delta_{t0}$. Let $\varphi = q$. We would have $\Phi(0, q) = 1$, but $\Phi(0, Pq) = \Phi(0, Fq) = 0$. So $\Phi(0, q \to Pq \wedge Fq) = 0$. □

This proposition undermines McTaggart's argument. Rostomyan then comes forward with the claim that McTaggart should have concluded that an event that is in the present, will be in the past and has been in the future. This claim is written as $\varphi \to PF\varphi \wedge FP\varphi$. He then shows that $\varphi \to PF\varphi \wedge FP\varphi$ is true in past and future serial flows of time. While he continues with commenting on McTaggart's response to the dual temporal (section 6.3.3), we leave his argument here. The rest is not of interest for the purpose of this thesis.

Not to bash on the author, but the paper's shortcomings illustrate nicely the nuances that come into play when using temporal logic as an analytical tool. While my final conclusion will turn out to be somewhat in agreement, I think that both the proof used and the conclusions miss the nuances. I will delve into a problem specific to Rostomyan's usage of temporal logic, and later I will discuss a more fundamental issue.

Let us begin with two small comments. In his paper, Rostomyan does not specify, but we now need to turn to the question of what events are. Since an event can be happening at any one moment, we need to have things that can be happening (true) or not happening (false) independently at each point. A good candidate would be the atomic propositions, as they can be either true or false at each of the moments,

with no inherent relation between the different points. Then in 6.1 we would not be talking about every formula φ but about every atomic proposition q.

This leads neatly into the second comment. The formula in 6.1 is a corollary of McTaggart's conclusion, rather than a phrasing in temporal logic. A better candidate would be the formula $\Box(Pq \wedge q \wedge Fq)$[22], where $\Box$ is the abbreviation introduced at the end of section 5.2.

Now, the main problem is that Rostomyan requires McTaggart's argument to hold for all of temporal logic, instead of just the class of flows of time which can describe our reality. This is overly restrictive, and does not mirror how modal logic is studied in general. It makes no sense to expect that the two requirements of irreflexivity and transitivity are sufficient in selecting flows of time that could describe reality. So showing that in the flow of time $(\mathbb{Z}, <)$ the formula $q \to Pq \wedge Fq$ is not necessarily true, does not undermine McTaggart's argument. Instead, it implies that $(\mathbb{Z}, <)$ is not an element of C_{True}, the class of flows of time that can describe our reality.

We can make this more precise using the ideas of axiomatization. The class C_{True} cannot be sound and complete with regard to the base system $\mathbf{B}$, since it is smaller then the class of all flows of time. But Rostomyan, by claiming that $\not\models q \to Pq \wedge Fq$ is sufficient to disprove McTaggart's argument, accidentally assumed that $\not\vdash_{\mathbf{B}} q \to Pq \wedge Fq$ is sufficient to disprove McTaggart. But that would only be the case if $\mathbf{B}$ was sound and complete to C_{True}.

To do this correctly, let us work with the assumption that there exists an axiomatization A_{True} which describes C_{True}[23]. Since it is the axiomatization of our universe, $\mathsf{C}_{\mathbf{B}+A_{True}}$ cannot be empty. We then come to my first main result:

Theorem 6.2. *The class $\mathsf{C}_{\mathbf{B}+A_{True}}$, which corresponds to the axiomatization of our universe, is empty if $\vdash_{\mathbf{B}+A_{True}} \Box(Pq \wedge q \wedge Fq)$.*

So under the assumption that the class $\mathsf{C}_{\mathbf{B}+A_{True}}$ is axiomizable in temporal logic, the condition of $\vdash_{\mathbf{B}+A_{True}} \Box(Pq \wedge q \wedge Fq)$ must hold if McTaggart's argument is right, since it follows from his argument on how our reality works.

Proof. Because of lemma 4.16.1, since $\vdash_{\mathbf{B}+A_{True}} \Box(Pq \wedge q \wedge Fq)$ we have that the two systems $\mathbf{B} + A_{True}$ and $\mathbf{B} + A_{True} + \Box(Pq \wedge q \wedge Fq)$ are equivalent, and hence sound and complete w.r.t. the same class $\mathsf{C}_{\mathbf{B}+A_{True}}$. From lemma 4.26, $\mathsf{C}_{\mathbf{B}+A_{True}}$ must then be a subclass of $\mathsf{C}_{\mathbf{B}+\Box(Pq \wedge q \wedge Fq)}$.

If we look at the class which is defined by $\Box(Pq \wedge q \wedge Fq)$, we can easily see that its empty. Indeed, it is clear that for every frame in every point where $\Box(Pq \wedge q \wedge Fq)$ is necessarily true, q must also be necessarily true. Any model that evaluates q as false in any point is thus a counterexample to $\mathcal{T} \models \Box(Pq \wedge q \wedge Fq)$.

Now we can use lemma 4.27 to show that $\mathsf{C}_{\mathbf{B}+\Box(Pq \wedge q \wedge Fq)}$ is empty. And since $\mathsf{C}_{\mathbf{B}+A_{True}}$ is a subset of $\mathsf{C}_{\mathbf{B}+\Box(Pq \wedge q \wedge Fq)}$, it must also be empty. □

Since the class $\mathsf{C}_{\mathbf{B}+A_{True}}$ is empty, the system $\mathbf{B} + A_{True}$ must be inconsistent if McTaggart's argument is true. So we have shown the incompatibility of temporal logic with McTaggart's argument. But much more then that we cannot say. That is the more fundamental issue with Rostomyan's argument. Since there is no proof that temporal logic is a correct lens through which we can examine reality, we cannot

[22]It is a good exercise to explain why and then to show that the upcoming proof also works with Rostomyan's formula: $q \to Pq \wedge Fq$.

[23]If it is not the case, then we can use that to argue that McTaggart's argument does not work well with Temporal Logic.

claim anything about the truth of McTaggart's argument, othen then that it is not compatible with temporal logic. And probably McTaggart would agree with this assessment, since he would likely reject the premises temporal logic are built upon. If one does not believe in time, as he did, one would say that only the now exist. So the flow of time that comes closest to fitting this world view is the one with only a single time point. But since that flow of time has an empty accessibility relation, we are effectively discarding every useful tool temporal logic gives us.

I want to end this section by describing my thoughts on the uses of temporal logic. These are not that well-founded, since I have only applied it to a single philosophical argument. But I do think there is value in sharing my thoughts on the subject. In using a tool, one is not required to accept its premises. But if one does not, the tool transforms into one to test the premises it is built upon.

If we have a possible collection of features of time, and we show that these features imply the truth of the premises, temporal logic can be used to explore what flows of time fit this collection of features. A possible result is that there are none. Then it has been mathematically proven that the collection of features cannot describe reality, since reality exists.

But if we do not know whether the features imply the premises, we can use temporal logic as a tool to prove it does not. This is what we have done with McTaggart's argument. But then one can only conclude that McTaggart collection of features he ascribes to time, cannot imply one of the premises of temporal logic. We cannot speak about the validity of McTaggart's argument.

With this, we leave the context of philosophy behind us, and I can lift the caveat mentioned at the beginning. Also for me, this was quite a different experience from what I normally encounter. At some points, I was afraid that there would not be enough mathematics in my thesis. Luckily, there is another aspect of the problem of time, which can be examined at the precision of pure mathematics.

7. Applying to Physics

At the start of last section, we mentioned that there are two quite different concepts underlying the problem of time. The philosopher's problem of time has already been discussed, so it is time to turn to what physicists call the problem of time. We will also quickly discuss a few different aspects of time that at first glance might seem ripe to be explored in the context of temporal logic, but which ran into fundamental issues. From this exploration one topic came forward, causal set theory. The rest of this section will then be focusing on the application of temporal logic to causal set theory.

7.1. **What is the Problem of Time?**

The physicist's problem of time refers to one specific issue within the larger problem of quantum gravity. Quantum gravity is one of the biggest issue for contemporary fundamental physics, i.e. the inability to unite general relativity with quantum field theory. General relativity is currently the best theory which describes the structure of the universe, but fails to predict anything on the sub-atomic quantum scale. Quantum field theory is the current theory which can predict what occurs on the quantum scales, and with a surprisingly good accuracy. But it fails when it tries to predict properties of the structure of the universe, like empty space.

The problem of time in physics is the recognition of one of the aspects where quantum mechanics and general relativity differ, namely their approach to time [14].

In quantum mechanics, time is usually taken as a background parameter, while in general relativity, it is part of spacetime, with a complex interaction with space and the presence of matter. And while quantum field theory is capable of working with the weaker theory of special relativity, it currently cannot work with the more powerful theory of general relativity.

The problem of time is a big topic, so the question is, to what part of it temporal logic applies in the most useful way. At first it was a question of difficulty. While quantum mechanics is something a bachelor physics student will be very familiar with, quantum field theory is more appropriate for a master's thesis. Similarly, most approaches to solving quantum mechanics are very involved. So first we started looking at how temporal logic could be used in general relativity, to gain a different view of how the theory could be understood and studied.

And the most promising seemed to be Lorentzian causality, the theory behind which spacetime points can have a causal effect on each other. But this theory uses topology, which cannot be described within temporal logic. It was during the study of Lorentzian causality that causal set theory came up, and this turned out to be the perfect candidate to apply temporal logic to.

7.2. **Causal Set Theory.** Over the years, there have been many approaches to quantum gravity. One of these approaches has a more mathematical angle: causal set theory. It postulates that the fundamental structure of the universe should not be described by a manifold, like general relativity proposes, but by a causal set.

Definition 7.1. A causal set C is a set with a relation $\prec$ that is:

- *Reflexive.* For all $x \in C : x \prec x$.
- *Transitive.* If $x_1 \prec x_2$ and $x_2 \prec x_3$ then $x_1 \prec x_3$.
- *Anti-symmetric.* If $x_1 \prec x_2$ and $x_2 \prec x_1$, then $x_1 = x_2$.
- *Finite intervals.* For all $x_1, x_2 \in C$, the interval
$$[x_1, x_2] := \{x \in C | x_1 \prec x \text{ and } x \prec x_2\}$$
is finite (and can possibly be empty).

The first three properties make the causal sets a partial order. The fourth property guarantees that locally the causal set is finite. But locally finite does not mean small, and the hypothesis is that the manifold general relativity prescribes to reality, is actually a high density causal set. The manifold is obtained by a continuum approximation, similar to how the finite molecules of a material are approximated as a solid, continuous object.

While we could talk about causal set theory for many more pages, I will instead suggest to read D.D. Reid *Introduction to causal sets: an alternate view of spacetime structure* [15] for an extended introduction which only requires passing knowledge of quantum mechanics and general relativity, and recommend S. Surya *The causal set approach to quantum gravity* [16] for a more complete understanding of the theory.

7.3. **Applying temporal logic.** At the moment, it might not be clear yet how temporal logic can be applied to causal set theory. The power of modal logic is to give a complete deduction system to many different classes of frames. And the causal sets, by the fact that they have relations, form a class of frames. But causal sets can more precisely be described as a well-defined class of flows of time, where the accessibility relation $<$ is derived from $\prec$ by excluding the reflexive $x \prec x$, i.e. $x < y$ if and only if $x \prec y$ and $x \neq y$. The question then becomes, if there is a system which

corresponds to the class of causal sets. And the first candidate is always to look at the defining formula. If φ defines the class of frames with the finite interval property FI, then there are good odds that $\mathbf{B} + \varphi$ is a system which is sound and complete to FI. Now we need to find a formula which defines the finite interval property.

7.4. **The Linear Case.** While researching the definability of the finite interval property, we do not have to start from scratch. There are already results in defining temporal logic *within the class of linear flows of time* (see section 5.5), for example Y. Vedema *Temporal Logic* [17]. So it is a good place to start looking into that first, and hopefully find a way to extend the scope to the general property of having finite intervals.

When looking at the class of linear flows of time K, Vedema offhand mentions the formula which defines the finite interval property within the class of linear frames. We need to prove it first to see how it works, and to that end we show a small lemma which will reduce our work.

Definition 7.2. For a flow $\mathcal{T} = (T, <)$, define the reverse flow $\mathcal{T}^\dagger = (T, <')$ over the same set T, where $t_1 <' t_2 \Leftrightarrow t_1 > t_2$.

It is clear that some properties, especially finite intervals, are unaffected when transforming $\mathcal{T}$ to $\mathcal{T}^\dagger$, while directional attributes will become their temporal conjugate. Now we can ask ourselves what happens to models of $\mathcal{T}$ after reversing:

Lemma 7.3. *If Φ is the extension for $(\mathcal{T}, \pi)$ defined in lemma 5.2, and $\Phi^\dagger$ is the extension for $(\mathcal{T}^\dagger, \pi)$, then $\Phi(t, \varphi) = \Phi^\dagger(t, \varphi^*)$.*

The proof is trivial once it is noted that in the definition for Φ in lemma 5.2 the semantic truth of $G\varphi$ is the order reverse of $H\varphi$, and the evaluation π only depends on the underlying set T.

Now we can introduce the defining formula.

Lemma 7.4. *The formula*

$$\varphi_{fi} = G(Gq \to q) \to (FGq \to Gq) \wedge H(Hq \to q) \to (PHq \to Hq)$$

defines the class of flows of time with finite intervals within the class of linear flows of time K.

We make two observations about φ_{fi}. First, φ_{fi} is temporal mirror invariant, i.e. $\varphi_{fi}^* = \varphi_{fi}$. Secondly, φ_{fi} can be split into two parts: φ_F, which only has future pointing modals, and φ_P, which has only past pointing modals. Then $\varphi_{fi} = \varphi_F \wedge \varphi_P$.

This proof is quite involved, but the fundamental idea behind it is straightforward. Examining only the φ_F half, we find that the $G(Gq \to q)$ guarantees us that, starting from a given point a, there are either zero or an infinite number of time points s with $\Phi(s, q) = 0$. The final Gq part makes φ_F only false in the infinite case. Then the FGq part enforces a later point b at which Gq is true. So we have an infinite series, which is bounded by a and b. Thus the interval $[a, b]$ must be infinite.

Proof. The proof goes via the contrapositive: A linear flow $\mathcal{T}$ doesn't have finite intervals if and only if there exists an evaluation π and a time point $t \in \mathcal{T}$ such that $\Phi(t, \varphi_{fi}) = 0$. We need to show the two implications.

The if: Assume $\mathcal{T}$ is a linear flow which doesn't have finite intervals. Then there must be an interval $[a, b] \subset \mathcal{T}$ which has infinite elements. For any infinite linear order, we can use the ascending descending sequence principle [18] to find a sequence

$\{t_n\}_{n\in\mathbb{N}}$, such that the sequence is either increasing or decreasing with respect to the ordering of $\mathcal{T}$.

If we only find a decreasing sequence, we can look at the sequence in $\mathcal{T}^\dagger$ where it will be ascending. The proof that follows will still be valid, since according to lemma 7.3 showing that φ_{fi} is false at a $t \in \mathcal{T}^\dagger$, will show that φ_{fi} is false at $t \in \mathcal{T}$.

Now, let us construct π as follows. For n even $\pi(t_n, q) = 0$, for n odd $\pi(t_n, q) = 1$, $\pi(b, q) = 1$ and for all other time points t, $\pi(t, q) = 1$. We will now show that $\Phi(t_0, \varphi_F) = 0$, which implies that $\Phi(t_0, \varphi_{fi}) = 0$.

Let $F = \{t \in \mathcal{T} | \forall_{n\in\mathbb{N}} t_n < t\}$, which is non-empty for it contains b. At each $t \in F : \Phi(t, Gq) = \Phi(t, q) = 1$. So there, $\Phi(t, Gq \to q) = 1$.

At any other time point $s \in \mathcal{T} \setminus F$, we can find the smallest even n such that $s < t_n$, and since $\Phi(t_n, q) = 0$, $\Phi(s, Gq) = 0$, which means that $\Phi(s, Gq \to q) = 1$.

Because for each $t \in \mathcal{T}$, $\Phi(t, Gq \to q) = 1$, it is also the case that at each time point $\Phi(t, G(Gq \to q)) = 1$.

Now if we look at t_0, we see that because $\Phi(b, Gq) = 1$, $\Phi(t_0, FGq) = 1$. But since $\Phi(t_2, q) = 0$, at t_0 we have $\Phi(t_0, Gq) = 0$. So $\Phi(t_0, FGq \to Gq) = 0$, which combined with $\Phi(t_0, G(Gq \to q)) = 1$ means that

$$\Phi(t_0, G(Gq \to q) \to FGq \to Gq) = \Phi(t_0, \varphi_F) = 0.$$

The only if: Assume that there is a linear flow of time $\mathcal{T}$, an evaluation π and a time point t such that $\Phi(t, \varphi_{fi}) = 0$. Since $\varphi_{fi} = \varphi_F \wedge \varphi_P$, we need to check both $\Phi(t, \varphi_F) = 0$ and $\Phi(t, \varphi_P) = 0$.

Assume $\Phi(t, \varphi_F) = 0$. We can rewrite $\Phi(t, \varphi_F) = 0$ as the formula

$$\Phi(t, G(Gq \to q) \wedge FGq \wedge \neg Gq) = 1,$$

by using the relation $\neg(A \to B) \leftrightarrow A \wedge \neg B$. Define $t_0 := t$. Because $\Phi(t_0, FGq) = 1$, there exists a $b > t_0$ such that $\Phi(b, Gq) = 1$. Now

$$\Phi(t_0, G(Gq \to q)) = 1 \Rightarrow \Phi(b, Gq \to q) = 1$$

Thus $\Phi(b, Gq) = \Phi(b, q) = 1$. Now, since $\Phi(t_0, \neg Gq) = 1$, there has to exist a $t_1 \in [t_0, b]$ such that $\Phi(t_1, q) = 0$.

Let us analyse $\Phi(t_1, \varphi_F)$. Since $\Phi(b, Gq) = 1$ we know that $\Phi(t_1, FGq) = 1$. And because $\Phi(t_0, G(Gq \to q)) = 1$ we have

$$\Phi(t_1, G(Gq \to q)) = \Phi(t_1, Gq \to q) = 1.$$

And since $\Phi(t_1, q) = 0$, it must be the case that $\Phi(t_1, Gq) = 0$. So

$$\Phi(t, G(Gq \to q) \wedge FGq \wedge \neg Gq) = 1.$$

Now we use the same reasoning for t_1 as for t_0 to conclude that there is a $t_2 \in [t_1, b]$ such that $\Phi(t_2, \varphi_F) = 0$. This process can be repeated indefinitely, giving us a sequence $\{t_n\}_{n\in\mathbb{N}} \subset [t_0, b]$. So the interval $[t_0, b]$ is infinite.

For φ_P, the same argument holds, changing the direction of the temporal aspects of the proof. So we can always conclude that $\mathcal{T}$ doesn't have the attribute of finite intervals. □

7.5. **The Non-Linear Case.** Having covered the linear case, we turn to the non-linear case. First, let us ask why the linear formula φ_{fi} fails. The problem comes down to the inability to guarantee that any two points have the property that one lays in the future of the other. And it is no surprise that that is the difficulty, since it is the fundamental difference between linear and non-linear partial orders.

Let us look at a concrete example. Let the flow of time consist of the set

$$T = \mathbb{N} + \mathbb{N} = \{(0, n) | n \in \mathbb{N}\} \cup \{(1, n) | n \in \mathbb{N}\},$$

and the accessibility relation $<$ be such that each subset $\mathbb{N}$ retains their inherited ordering, but with $(0, 0) < (1, 0)$ taking the transitive closure, i.e. adding all relations $x < y$ such that $<$ becomes transitive. We are left with two lines of natural numbers, which only overlap in their starting point. It is clear that this frame does not have any infinite intervals.

Now let the evaluation be $\pi((i, n), q) = \delta_{i0}$, meaning that along one line and the shared point, q evaluates to 1, and along the other line q evaluates to true. Now, checking all options, it is trivial to see that $\Phi(t, G(Gq \to q)) = 1$. And at all t one has $\Phi(t, FGq \to Gq) = 1$, except for $t = (0, 0)$. There, $\Phi((0, 0), FGq) = 1$, but $\Phi((0, 0), Gq) = 0$. This is possible since the future point that makes FGq true has nothing in common with the time point that makes Gq false.

One could attempt to fix this underlying issue, and one such approach we will explore. Using $G(Gq \to q)$ and $\neg Gq$, we are still required to have an infinite series of points where q is false. But the FGq part was not sufficient in forcing this infinite series to be contained within an interval, since FGq requires only one future to have Gq. What if we replaced FGq with $G(\neg q \to FGq)$? This would enforce every point in the infinite series to have a future where Gq. The formula would then become

$$G(Gq \to q) \to (G(\neg q \to FGq) \to Gq).$$

This would be true at $t = (0, 0)$ in the last example, but does not solve the problem.

We can just as easily find a counterexample to this proposed defining formula. Let $T = \mathbb{N} \times \mathbb{N}$, where $(a, b) \leq (x, y)$ if $a \leq x$ and $b \leq y$, from which $<$ has it's natural meaning. Then define $\pi((0, n), q) = 0$ for all $n \in \mathbb{N}$, and for all other (x, y), define $\pi((x, y), q) = 1$.

We have again that $\Phi(t, Gq \to q) = 1$ for all $t \in T$, and $G(\neg q \to FGq)$ is also true everywhere. For every $(x, y) \in T$ with $x > 0$, there are no later s with $\Phi(s, \neg q) = 1$. So for all (x, y) with $x > 0$: $\Phi((x, y), G(\neg q \to FGq)) = 1$. Then for every $(0, y) \in T$, it is also the case that $\Phi((0, y), \neg q \to FGq) = 1$, but there it is also the case that $\Phi((0, y), Gq) = 0$, making the new formula false in this situation which again has no infinite intervals.

While I did study other possible alternatives, they all had obvious counterexamples. So the obvious next step is to prove that it is impossible to define the finite interval property. And that can be done, but only by looking at a new way one can have an infinite interval in the non-linear case, which is impossible in linear case.

7.6. **Finite Chains.** So far, we have examined the finite interval property in non-linear flows of time in the same way we examined the property in linear flow of time. The infinite intervals all contained a sub-chain that is an infinite interval. But there is a unique way in which a non-linear ordering can have an infinite number of elements bounded by two points a and b. Instead of a singular chain that has an infinite number of elements, we will be looking at an infinite number of chains between a and b. This way the interval is infinite, while each path taken from a to b is finite.

What we are going to show, is that there cannot be a formula that defines the property of having a finite number of chains between any two points. This is done by once again using bounded morphisms (definition 5.9).

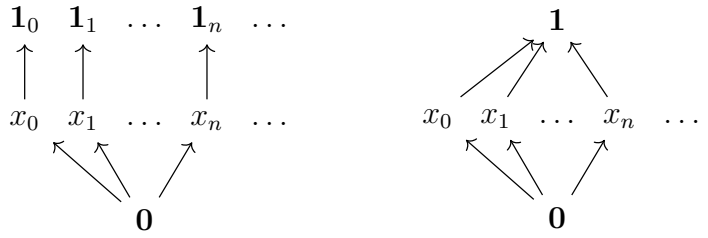


FIGURE 2. Left: The causal set $\mathcal{T}_{fi}$ has only finite intervals. Right: The flow of time $\mathcal{T}_\infty$ has one infinite interval $[\mathbf{0}, \mathbf{1}]$.

The definition of bounded morphisms can be extended to frames in general, instead of proper models as in definition 5.9. This is done by dropping the model preserving property. We are then left with the following definition of bounded morphisms between frames.

Definition 7.5. A bounded morphism is a mapping $(T, <) \mapsto (T', <')$ defined by its derived mapping, $T \to T'$ such that:

- $t < s$ implies that $t' <' s'$. (Forwards preserving)
- $t' < x$, where $x \in T'$ implies that there is an $s \in T$ with $t < s$ and $s' = x$. (Backwards consistency)

With this definition, Blackburn *et al. Modal Logic* [7] Theorem 3.14.iii states that:

Lemma 7.6. *If there exists a bounded morphism $\mathcal{T} \mapsto \mathcal{T}'$ and $\mathcal{T} \models \varphi$ then $\mathcal{T}' \models \varphi$.*

Now let us examine the two specific frames shown in figure 2. First we have $\mathcal{T}_{fi}$, which only has finite intervals and is thus a causal. The set $T_{fi} = \bigcup_{i\in\mathbb{N}}\{x_i, \mathbf{1}_i\} \cup \{\mathbf{0}\}$ has the relation $<_{fi}$ such that $\mathbf{0} <_{fi} x_i$, $\mathbf{0} <_{fi} \mathbf{1}_i$ and $x_i <_{fi} \mathbf{1}_i$. This result in a frame with a first point, $\mathbf{0}$, from which an infinite number of chains start. But each of the chains goes to a unique end point, making the interval $[\mathbf{0}, \mathbf{1}_i] = \{\mathbf{0}, x_i, \mathbf{1}_i\}$, which is clearly finite. So the frame has only finite intervals.

The second frame will be $\mathcal{T}_\infty$, which will have an infinite interval and is thus not a causal set. The set $T_\infty = \bigcup_{i\in\mathbb{N}}\{x_i\} \cup \{\mathbf{0}, \mathbf{1}\}$ has the relation $<_\infty$ such that $\mathbf{0} <_\infty \mathbf{1}$ and for every $i \in \mathbb{N}$ $\mathbf{0} <_\infty x_i$ and $x_i <_\infty \mathbf{1}$. This frame consists of a first point, $\mathbf{0}$, and a final point, $\mathbf{1}$, between wich there are an infinite number of chains. So the interval $[\mathbf{0}, \mathbf{1}]$ is infinite.

We can then define a bounded morphism $\mathcal{T}_{fi} \mapsto \mathcal{T}_\infty$ via the derived mappings $T_{fi} \mapsto T_\infty$, $\mathbf{0} \mapsto \mathbf{0}$, $x_i \mapsto x_i$, $\mathbf{1}_i \mapsto \mathbf{1}$ and $<_{fi} \mapsto <_\infty$. We need to check the two properties from definition 7.5.

First, let us check the forwards preserving property. We need to check the three possibilities separately. Because $\mathbf{0} <_{fi} x_i$ it must be the case that $\mathbf{0} <_\infty x_i$. Because $\mathbf{0} <_{fi} \mathbf{1}_i$ it must be the case that $\mathbf{0} <_\infty \mathbf{1}$. And because $x_i <_{fi} \mathbf{1}_i$ it must be the case that $x_i <_\infty \mathbf{1}$. All three are indeed true, meaning that our mapping has the forwards preserving property.

Secondly, we check the backwards consistency property. This time there are two possibilities. Let $s' \in T_\infty$, and have $\mathbf{0} <_\infty s'$. Whether s is an x_i or $\mathbf{1}$, there is an $s \in T_{fi}$ such that $\mathbf{0} <_{fi} s$ and $s \mapsto s'$. If we instead have $x_i <_\infty s'$, then $s' = \mathbf{1}$, and then there is $\mathbf{1}_i \in T_{fi}$ with $x_i <_f i\mathbf{1}_i$ and $\mathbf{1}_i \mapsto \mathbf{1}$.

With these concepts introduced, we can show the second main result of this thesis.

Theorem 7.7. *There is no formula φ_{fi} that defines the class of flows of time with the finite interval property.*

Proof. Let φ_{fi} be a formula that defines the class of flows of time with the finite interval property. Then $\mathcal{T}_{fi}$ must be an element of the class, since it is a flow of time with the finite interval property. This means that $\mathcal{T}_{fi} \models \varphi_{fi}$. But since there is a bounded morphism $\mathcal{T}_{fi} \mapsto \mathcal{T}_{\infty}$, lemma 7.6 tells us that $\mathcal{T}_{fi} \models \varphi_{fi}$ implies that $\mathcal{T}_{\infty} \models \varphi_{fi}$. But then by definition 4.11, this requires that $\mathcal{T}_{\infty}$ is part of the class of flows of time with the finite interval property. This contradiction means that there cannot be any formula φ that defines the finite interval property. □

With this theorem, we know that the simple system $\mathbf{B} + \varphi$ cannot be sound and complete with respect to the class of causal sets. So if we have causal sets as described in definition 7.1, there needs to be a creative usage of axioms to find an appropriate system. Alternatively, the definition of causal sets could be sharpened, as not all causal sets are actually of interest for causal set theory. But both these possibilities are left as questions for further research.

References

[1] Nicolaas P. Landsman. Propositielogica: Onderdeel van het college logica (2017), `https://www.math.ru.nl/~landsman/Propositielogica2017.pdf`, 2017.

[2] Ieke Moerdijk and Jaap van Oosten. *Sets, models and proofs*. Springer, 2018.

[3] James Garson. Modal Logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. `https://plato.stanford.edu/archives/fall2018/entries/logic-modal/`, fall 2018 edition, 2018.

[4] Nino B Cocchiarella. Notes on deontic logic. *https://www.ontology.co/essays/deontic-logic.pdf*, 2015.

[5] Daniel Rönnedal. Doxastic logic: a new approach. *Journal of Applied Non-Classical Logics*, 28(4):313–347, 2018.

[6] Johan Van Benthem. *Modal logic for open minds*. Center for the Study of Language and Information Stanford, 2010.

[7] Patrick Blackburn, Maarten De Rijke, and Yde Venema. *Modal logic*, volume 53. Cambridge University Press, 2002.

[8] Barry Dainton. *Time and space*. Routledge, 2016.

[9] Craig Callender. *What makes time special?* Oxford University Press, 2017.

[10] Jc Beall, Michael Glanzberg, and David Ripley. Liar paradox. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. `https://plato.stanford.edu/archives/fall2020/entries/liar-paradox/`, fall 2020 edition, 2020.

[11] J Ellis McTaggart. The unreality of time. *Mind*, 17(68):457–474, 1908.

[12] EJ Lowe. McTaggart's paradox revisited. *Mind*, 101(402):323–326, 1992.

[13] Hunan Rostomyan. McTaggart's argument for the unreality of time: A temporal logical analysis. *Harvest Moon: New Crop Prize Edition (UC Berkeley's Undergraduate Journal of Philosophy)*, 3, `https://www.researchgate.net/publication/258180410_McTaggart%27s_Argument_for_the_Unreality_of_Time_A_Temporal_Logical_Analysis` 2014.

[14] Edward Anderson. Problem of time in quantum gravity. *Annalen der Physik*, 524(12):757–786, 2012.

[15] David D Reid. Introduction to causal sets: an alternate view of spacetime structure. *arXiv preprint gr-qc/9909075*, 1999.

[16] Sumati Surya. The causal set approach to quantum gravity. *Living Reviews in Relativity*, 22(1):5, 2019.

[17] Yde Venema. Temporal logic. *The Blackwell Guide to Philosophical Logic*, 1998.

[18] Denis R Hirschfeldt and Richard A Shore. Combinatorial principles weaker than ramsey's theorem for pairs. *The Journal of Symbolic Logic*, 72(1):171–206, 2007.