Axioms of deliberative *stit* inspected

A bachelor thesis by R W Seń
Student number: 3472361
Supervising professor: Dr. J. Broersen

Summer 2013

Abstract

This bachelor thesis offers a basic introduction to stit logic, before looking into two possible problems with [1] noted by my supervisor Dr. Jan Broersen.

Keywords: stit, axioms

1 Introduction

Stit logic is a logic used to describe the concurrent actions that affect the world performed by a group of agents. Stit stands for *seeing to it that*, meaning that agent has done the specified action. In other words, that agent is responsible for that action.

The theory of stit is interesting in the field of AI, because it offers an alternative logic for groups of multiple agents. It differs from other multi-agent logics in the way time is approached: time is less a sequence of states, and more a continuous time line with certain relevant moments specific to each agent. This approach resonates better with the intuitive human understanding of time.

In this paper I will explain the basics of stit logic, covering the necessary background of Branching Time in the process.

After laying the background theory, I will investigate two specific questions asked by my supervisor, Jan Boersen, about the proofs for the alternative axioma's in [1]. These questions are:

1. In the proof for lemma 2 of [1], is the step from line 5 to line 6 correct? — Intuively, a distribution of the $\Diamond$ operator is not staightforward.

2. How does the expansion of the AAIA axioma from the two-agent case to the three-agent case ensure Independence of Agents? — The exact formulation of the AAIA axioma is quite short, and it is not immidiately obvious how this axiom is supposed to work.

I will deal mainly with the deliberate version of stit logic and when mentioning stit logics other than deliberative, I will explicitly say so.

I recommend that readers new to stit logic read the first three sections from [1] after reading my introduction to stit logic, but before I investigate the above problems.

2 What is stit logic?

Stit logic is a modal extension of proposition logic where the added operator $[i\ stit : \varphi]$ accounts for the agency of agent i in φ, meaning agent i has seen to it that φ. Stit theories are set in a Branching Time theory, an non-deterministic temporal framework first proposed by [4], and exhaustively explained in [3]. I will cover Branching Time theory summarily, before moving on to stit itself.

2.1 A branching tree theory of time

Branching Time (BT) is a structured way for looking at time, providing the necessary abstraction to do so logically. The intuitive idea is as follows:

consider a segment of time as a line. Spread across this line are moments: certain instances of time where something interesting happens. At every moment there is an non-deterministic event which splits the timeline into two or more possible continuations or possible futures, transforming the line into a tree with many branches. In stit theories, these events represent moments in time whereat an agent can make a choice.

Branching Time consists of moments ordered in a forward branching treelike structure, "with forward branching representing the openness or indeterminacy of the future, and the absence of backward branching representing the determinacy of the past"[3].

This is formally represented as the tuple $\langle Moments, < \rangle$ where $Moments$ is the non-empty set of moments and $<$ is a tree-like ordering on $Moments$, such that for any m_1, m_2 and m_3 in $Moments$ if $m_1 < m_3$ and $m_2 < m_3$, then either $m_1 = m_2$ or $m_1 < m_2$ or $m_1 > m_2$[3].

Every maximal set of linearly ordered moments is called a history, with $m \in h$ meaning that moment m is in history h. Histories are essentially possible timelines, while moments are points in time where two or more histories are differentiated. All histories that run through a certain moment m constitute the set $H(m) = \{h : m \in h\}$.

2.2 Stit

In order to understand models of stit, it is necessary to define the concepts *Agents*, *Choices* and *Instants*.

2.2.1 Agents, Choices and Instants

Stit deals with agency in logic and for there to be agency, there has to be an agent. In any model of stit logic, a set AGT of agents is defined. These agents are individuals that make choices influencing the world around them.

Next we introduce the concept of choices. At each moment m of our model one or more agents can make a choice. An agent can make a choice if it can constrain the possible future histories to a subset of all histories going through m. In other words: an agents makes a choice by making certain future histories impossible, whilst ensuring there is at least one possible future history.

Formally, these choices are collected in the Choice function.

$$Choice : AGT \times Moments \rightarrow 2^{2^{Hist}}$$

is a function where $Hist$ is the set of all histories. This function maps each agent and each moment onto a partition of $H(m)$ [1]. See figure 1.

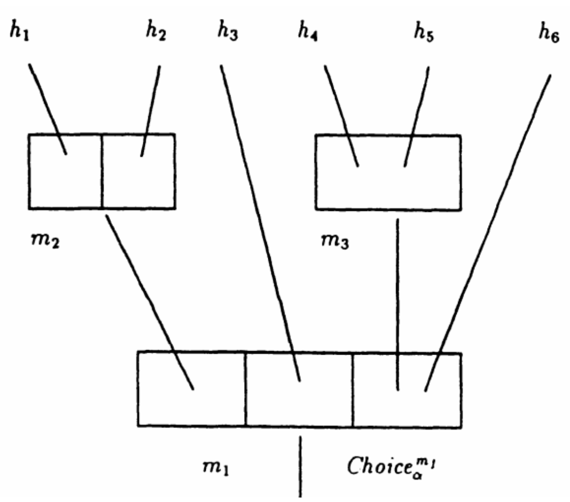


Figure 1: A BT tree containing three moments and six histories. h_4, h_5 and h_6 are $Choice_\alpha^{m1}$-equivalent, meaning that they fall in the same choice partition for agent α at moment m_1. Source: [3]

Finally we introduce the concept of instants. Instants are collections of moments from different histories that can be thought of as occuring at the same time. Moments from the same instant are therefore temporal alternatives to eachother. Instances are used to compare alternative histories at certain moments.

Formally, all instants are collected in the set $Instant$, defined as the set of instants partitioning the moments of $Moments$ horizontally into equivalence classes [3]. The instant to which a certain moment m belongs is given by $i_{(m)}$.

2.2.2 Choice-equivalence

We have established an intuitive framework of time wherein agents can make choices. In order to reason about these choices we introduce proposition logic and extend that with the stit operator: $[i\ stit : \varphi]$.

Before we turn to the definition of the stit operator, we will define the auxilary concept of $Choice_i^m$-equivalence. Suppose that an agent α has a choice with three options at moment m_1, as seen in figure 1. He can make a choice, ensuring that either 1) h_1 or h_2 is the future; 2) h_3 is the future; or 3) either h_4, h_5 or h_6 is the future. The agent can not distinguish between histories within a $Choice_\alpha^{m_1}$-partition, because if there is a point at which two histories differ at moment m_1 from the perspective of the agent, said agent could choose between them, neseccitating another choice-partition.

We capture this concept formally with with $Choice_i^m$-equivalence. Sup-

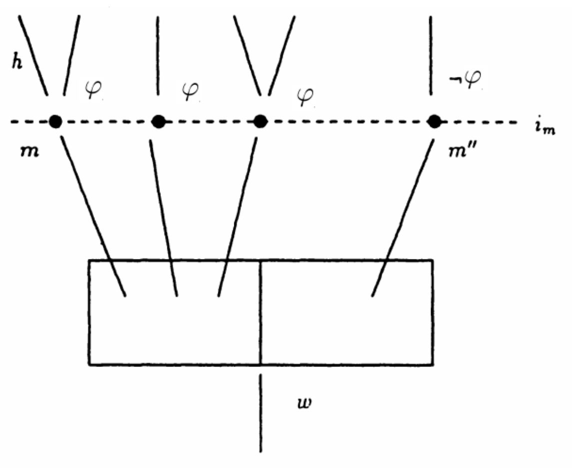


Figure 2: An example of a more complete stit model. Note that the agent can not distinguish between choice-equivalent histories, and that m and m'' are part of the same instant. Source: [3]

pose that moments m_1 and m_2 occur at the same instant, so $i_{(m_1)} = i_{(m_2)}$. If m_1 and m_2 fall within the same $Choice_i^{m_1}$-partition, they are said to be $Choice_i^{m_1}$-equivalent.

2.3 The stit operator

The operator $[i\ stit : \varphi]$ tells us that agent i is agentive in φ, in other words i sees to it that φ is true. Stit models are an extension of Branching Time models, and thus every $[i\ stit : \varphi]$ has to happen at a certain moment m. $[i\ stit : \varphi]$ can be understood as "agent i sees to it that the future history is contained within those future histories where φ is true". In other words: the agent removes possible future histories where $\neg\varphi$ holds. See for example figure 2. On top of the guarantee that φ is true, it is also necessary that on at least one alternative to a future moment[1] φ is false. If not, how can an agent claim to have seen to it that φ if φ would have been the case in every possible future?

2.3.1 Historical necessity

In stit models, situations occur where in every possible future, a certain formula φ is true. Here it can not be said that one of the agents is responsible, because there is no alternative future history where $\neg\varphi$ holds.

When these situations occur, it is said that φ is *historically necessary.* This is logically represented by the operator $\Box$. Dual to formulas of the form

[1]This alternative moment comes from a history not in the set of future histories, but it is in the same instant as a moment in the set of future histories.

$\Box\varphi$ is the operator for *historical possibility*: $\Diamond\varphi$. It follows the standard definition $\Diamond\varphi =_{def} \neg\Box\neg\varphi$.

2.3.2 Future and Past operators

In stit models, it is necessery to reference past and future states. The stit language incorporates the operators $F\varphi$ and $P\varphi$, for future and past respectively, for just this purpose. Fφ and Pφ follow conventional interpertations, and will be fully defined in the coming paragraphs.

2.3.3 Languages of stit

A stit-language $\mathcal{L}^{AGT}$ is now defined by the following Backus Naur Form:

$$\varphi ::= p|\neg\varphi|(\varphi \wedge \varphi)|[i\ stit : \varphi]|\Box\varphi|\mathrm{F}\varphi|\mathrm{P}\varphi$$

where p is an atomic proposition and $i \in AGT$.

Models of $\mathcal{L}$ are defined as a tuple $\mathcal{M} = \langle Moments, <, Choice, V\rangle$, where $\langle Moments, <\rangle$ is a BT structure, $Choice$ is the above defined choice-function, and V is a valuation function $V : ATM \rightarrow 2^{Moments}$ where ATM is the set of atomic propositions and $Moments$ is the set of all moments [1].

2.4 Truth conditions in stit models

To round out this introduction to stit I present the semantics of a stit formula, even though we do not concern ourself with semantics in the core section. The truth value of a formula in a stit model is calculated as follows:

- $\mathcal{M}, m/h \models A$ iff $m/h \in V(A)$ for A an atomic formula,
- $\mathcal{M}, m/h \models A \wedge B$ iff $\mathcal{M}, m/h \models A$ and $\mathcal{M}, m/h \models B$,
- $\mathcal{M}, m/h \models \neg A$ iff $\mathcal{M}, m/h \not\models A$,
- $\mathcal{M}, m/h \models \Box A$ iff $\mathcal{M}, m/h' \models A$ for all $h' \in H_{(m)}$,
- $\mathcal{M}, m/h \models \mathrm{P}A$ iff there is an $m' \in h$ such that $m' < m$ and $\mathcal{M}, m'/h \models A$,
- $\mathcal{M}, m/h \models \mathrm{F}A$ iff there is an $m' \in h$ such that $m < m'$ and $\mathcal{M}, m'/h \models A$,
- $\mathcal{M}, m/h \models [i\ stit : \phi]$ iff (1) $M, m/h \models A$ for each $h' \in Choice_i^m(h)$, and (2) there is some $h'' \in H_{(m)}$ for which $\mathcal{M}, m/h \not\models A$.

As found in [3].

3 Possible problems with the Alternative Axioma for Independence of Agents in [1]

In this section, I will examine the two main research questions of this paper:

1. How does the expansion of the AAIA axioma from the two-agent case to the three-agent case ensure Independence of Agents?
2. In the proof for lemma 2 of [1], is the step from line 5 to line 6 correct?

However, before doing this, I will define the subdomain in which these problems take place.

As this is an inspection of certain aspects of [?], I highly recommend reading the first three sections of that paper. Otherwise, the following might not make sense.

3.1 Subdomain

The specific problems I will deal with all concern logical constructs within an arbitrary moment. Therefore, the operators F and P are not utilized.

3.2 Possible problem with the interpretation of three agent case of AAIA

The axiomatics of [1] are build upon those of Xu[5]. "Xu gave the following axiomatics of Chellass STIT:

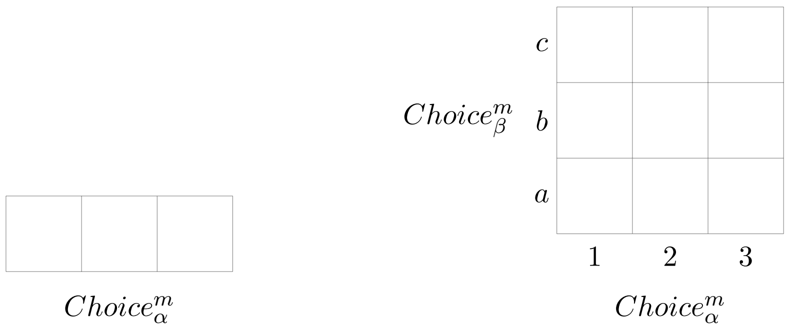
- S5($\Box$) — the axiom schemas of S5 for $\Box$
- S5(i) — the axiom schemas of S5 for every $[i\ stit : \varphi]$.
- $(\Box \to i)$ — $\Box\varphi \to [i\ stit : \varphi]$
- (AIA_k) — $(\Diamond[0]_{\varphi 0} \wedge \cdots \wedge \Diamond[k]_{\varphi k}) \to \Diamond([0]_{\varphi 0} \wedge \cdots \wedge [k]_{\varphi k})$

The last item is a family of axiom schemes for independence of agents that is parameterized by the integer k." These axioms are interpreted for $k+1$ agents and $[i\ stit : \varphi]$ is shortened to $[i]\varphi$[2].

Balbiani et al. replace Xu's *Axiom schema for Independence of Agents* (AIA) with their own version: the *Alternative Axiom schema for Independence of Agents* (AAIA).

- $(AAIA_k)$ — $\Diamond\varphi \to \langle k\rangle \bigwedge_{0\leq i<k}\langle i\rangle\varphi$ for $k \geq 1$

[2]The dual operator to $[i\ stit : \varphi]$: $\langle i\ stit : \varphi\rangle =_{def} \neg[i\ stit : \neg\varphi]$ is shortened as $\langle i\rangle\varphi$.



(a) Choice space for 1 agent with three options.

(b) Choice space for 2 agents, each with three options.

Figure 3: Examples of choice space for one and two agents.

This, too, is a family of axiom schemes, for $k + 1$ agents. However, it is not exactly clear how this translates into usable logic. For example, while Xu's (AIA_{k+1}) implies (AIA_k), Balbiani et al.'s $(AAIA_{k+1})$ does not imply $(AAIA_k)$.

I will first demonstrate what *choice space* is and what it looks like when choices are independent. Then I will demonstrate how $(AAIA_k)$ fulfills independence of agents.

3.2.1 Visualizing choices: choice space

This method of visualizing choices assists understanding of the concept independence of agents. This method is used in [3], but it is given no name. I will call this method *choice space*, because the visualizing of choices with this method depends on the visualizing of n-dimensional spaces.

See figure 3a for an example of the choice space for one agent α with three options. This choice is called $Choice^m_\alpha$, because it is represents the options agent α has in moment m.

Next we visualize the choice space for two agents. See figure 3b for an example of the choice space where two agents each have three options. The options for each agent are labeled, to facilitate discussion of the choice space.

The dimensionality of choice spaces is equal to the number of agents that have a choice at that moment. So a three agent choice space would have three dimensions, and so forth.

Finally, figure 4 is an example of 2 agents that have the ability to restrict each others options. Here it is not possible for agent α to choose 3 and for agent β to simultaniously choose c.

It now is easy to see if a particular choice space satisfies independence of agents (IA): a two-agent choice space satisfies IA if the choice space is a rectangle; a three-agent choice space satisfies IA is the choice space is a cuboid; *et cetera*. We will call choice spaces satisfying IA *complete*.

To see why it is important that stit satisfies independence of agents at all, remember that choice are made in moments. These moments are timeless, i.e. time only progresses between moments, not during a moment itself. If the agents from example 4 were to make their choices serially, there would be no problem in one agent restricting the options of the other agent. However, agents α and β make their choices concurrently. If it were possible to have choice spaces that do not satisfy independence of agents, it would be possible to have agents choose in such a way that there is no future.

This is a rather meaningless situation, so therefore *Independence of Agents* is fundamental to stit.

3.2.2 How Balbiani et al.'s $(AAIA_k)$ satisfies independence of agents

To see how Balbiani et al.'s $(AAIA_k)$ satisfies indepence of agents, here are some examples. I will not prove the correctness of $(AAIA_k)$ for any k, I will only show how $(AAIA_k)$ generates complete choice spaces.

The **two agent case**, where $k = 1$, looks like this:
$(AAIA_1)$: $\Diamond\varphi \rightarrow \langle 1\rangle\langle 0\rangle$
Translated to english, this means that if φ is possible, then agent 1 possibly sees to it that agent 0 possibly sees to it that φ is true.

In other words, It is impossible for agent 1 to make such a choice that something that was possible becomes impossible for agent 0, thereby garanteeing a rectangular choice space.

The **three agent case**, where $k = 2$, looks like this:

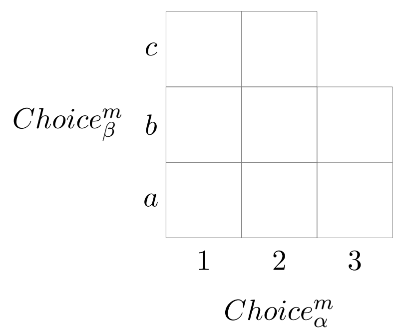


Figure 4: Choice space for 2 agents, each with three options.

$(AAIA_2)$: $\Diamond\varphi \to \langle 2\rangle(\langle 0\rangle\varphi \wedge \langle 1\rangle\varphi)$
Translated to english, this means that if φ is possible, then agent 2 possibly sees to it that agents 0 and 1 possibly see to it that φ is true.

In other words, agent 2 cannot choose in a way that constricts the choices of either agent 0 or 1. This in and of itself is not enough to satisfy independence of agents, since agent 1 could now make a choice that eliminates options for agent 0.

But [1] solved this problem. $(AAIA_k)$ is not simply a variable axiom schema dependent on the number of agents, it is a *family of axiom schemes* dependent on the number of agents. So, when looking at the three agent case, not only do we have the axiom $(AAIA_2)$: $\Diamond\varphi \to \langle 2\rangle(\langle 0\rangle\varphi \wedge \langle 1\rangle\varphi)$, but we also have $(AAIA_1)$: $\Diamond\varphi \to \langle 1\rangle\langle 0\rangle$! The phrasing that hints at this interpertation is rather awkward, but the intention becomes clearer when the wording of this concept in [**?**] is taken into account. Xu defines his families more explicit, where Balbiani et al./ thought the meaning of *family of axiom schemas* apparent.

This means that agent 2 can not make a choice that limits any agent in its choices (every slice of the choice space is equally shaped), and the choice space for the other two agents is a rectangle, courtesy of $(AAIA_1)$. This guarantees that the choice space, now shaped as a stack of rectangles, is a cuboid, and thus complete.

By this method, all higher dimensional choice spaces are contructed: $(AAIA_k)$ states that the choice spaces of dimension $k-1$ are equally shaped, and $(AAIA_{k-1})$ states that all choice spaces of dimension $k-1$ satisfy IA.

3.3 Possible problem in lemma 2

In [1], Balbiani et al. prove lemma 2, a lemma that states that Xu's Axiom schema for the Independence of Agents (AIA_k)[5] follows from their own axiom schema $AAIA_k$, and the schemas $S5(\Box)$, $S5(i)$ and $(\Box \to i)$.

The proof of lemma 2 is fairly straightforward, except for the step from line five to line six. In this step, a $\Diamond$ is distributed to a lower level in the formula. This seems counter-intuitive. The the exact step in the proof is:
Line 5:

$$\Diamond(\Diamond[0]\varphi_0 \wedge [1]\varphi_1) \to \Diamond\langle 1\rangle([0]\varphi_0 \wedge [1]\varphi_1)$$

is equivalent to line 6:

$$\Diamond[0]\varphi_0 \wedge \Diamond[1]\varphi_1 \to \Diamond\langle 1\rangle([0]\varphi_0 \wedge [1]\varphi_1)$$

Since these are rather cumbersome formula's, I will use these notational shorthands:
$A = [0]\phi_0$

$B = [1]\phi_1$
$C = \Diamond\langle 1\rangle(A \wedge B)$

To prove that this is a correct step, I will show the antecedents of both formulas to be equivalent. This is enough for equivalence, since the consequents of the formula's are identical.

$$\Diamond(\Diamond A \wedge B) \leftrightarrow (\Diamond A \wedge \Diamond B)$$

I will only need the schema $S5$ of the $\Box$ and $\Diamond$ operators for the following proofs. I will split the proof in two seperate parts. The proofs are in the fitch-style notation, introduced in [2].

3.3.1 The $\rightarrow$ direction

To prove: $\Diamond(\Diamond A \wedge B) \rightarrow \Diamond A \wedge \Diamond B$

1	$\Diamond(\Diamond A \wedge B)$	Assumption
2	$\quad \Box$	$\Box$
3	$\quad \Diamond A \wedge B$	Assumption
4	$\quad B$	Elim $\wedge$, from line 3
5	$\Diamond B$	Intro $\Diamond$, from line 2,4
6	$\quad \Box$	$\Box$
7	$\quad \Diamond A \wedge B$	Assumption
8	$\quad \Diamond A$	Elim $\wedge$, from line 7
9	$\Diamond\Diamond A$	Intro $\Diamond$, from line 2,8
10	$\Diamond A$	Theorem $4\Diamond$[3], from line 9
11	$\Diamond A \wedge \Diamond B$	Intro $\wedge$, from line 6,10
12	$\Diamond(\Diamond A \wedge B) \rightarrow \Diamond A \wedge \Diamond B$	Intro $\rightarrow$, from line 1,11

3.3.2 The $\leftarrow$ direction

To prove: $(\Diamond A \wedge \Diamond B) \rightarrow \Diamond(\Diamond A \wedge B)$

[3] $4\Diamond$ is the dual theorem to the regular theorem 4 (transitivity) in S5. $4\Diamond$: $\Diamond\Diamond\varphi \rightarrow \Diamond\varphi$

1	$\Diamond A \wedge \Diamond B$	Assumption
2	$\Diamond A$	Elim $\wedge$, from line 1
3	$\Box\Diamond A$	axiom 5, from line 2
4	$\Diamond B$	Elim $\wedge$, from line 1
5	$\Box$	$\Box$
6	B	Assumption
7	$\Diamond A$	Elim $\Box$, from line 3
8	$\Diamond A \wedge B$	Intro $\wedge$, from line 6,7
9	$\Diamond(\Diamond A \wedge B)$	Intro $\Diamond$, from line 8
10	$(\Diamond A \wedge \Diamond B) \rightarrow \Diamond(\Diamond A \wedge B)$	Intro $\rightarrow$, from line 1, 9

3.3.3 Equality

As shown in the two proofs above, $\Diamond(\Diamond A \wedge B) \leftrightarrow (\Diamond A \wedge \Diamond B)$ is correct, and therefore $(\Diamond(\Diamond A \wedge B) \rightarrow C) \leftrightarrow ((\Diamond A \wedge \Diamond B) \rightarrow C)$ is correct.

This proves the step from line 5 to line 6 in lemma 2 of [1] is correct.

4 Conclusion

4.1 Concluding remarks on axioms for stit

The questions I have investigated in this paper are:

1. In the proof for lemma 2 of [1], is the step from line 5 to line 6 correct?
2. How does the expansion of the AAIA axioma from the two-agent case to the three-agent case ensure Independence of Agents?

I have answered these questions as follows:

1. This step is indeed correct. See for the proof.
2. The AAIA axiom is defined as a "family of axiom schemes". This formulation is not as easily understandable as wanted, instead trusting on the similarities between this axioma and Xu's axioma in [5] for the reader to figure it out. When interpreting the AAIA axiom as intended, the results are valid.

I conclude there is no fault in these two parts of [1].

4.2 Further investigation

The following subjects are interesting for further studies:

- producing a working implementation of stit logic;
- transforming the above set of axioms for multi-agent stit into a group or multi-group set of axioms;
- investigating the possibility to use stit logic as a logical model for responsibility.

References

[1] Philippe Balbiani, Andreas Herzig, and Nicolas Troquard. Alternative axiomatics and complexity of deliberative stit theories. *Journal of Philosophical Logic*, 37(4):387–406, 2008.

[2] Jon Barwise, John Etchemendy, Gerard Allwein, Dave Barker-Plummer, and Albert Liu. *Language, proof and logic*. CSLI publications, 2000.

[3] John F Horty and Nuel Belnap. The deliberative stit: A study of action, omission, ability, and obligation. *Journal of philosophical logic*, 24(6):583–644, 1995.

[4] Arthur N Prior. Past, present and future. 1967.

[5] Ming Xu. Axioms for deliberative stit. *Journal of Philosophical Logic*, 27(5):505–552, 1998.