Lecture Notes 10 – Solutions to the Frame Problem(s) I

Patrick Doherty

February 7, 1998

1 Outline

1. Yale Shooting Problem
2. Initial Solutions
   (a) Chronological Minimization
   (b) Causal Minimization
   (c) Explanation Closure
3. Chronological Minimization
   (a) Lifschitz's Work
   (b) Kautz's Work
   (c) Shoham's Work

2 Introduction

In the previous lectures, we learned about logic-based approaches to KR and were introduced to problems which arise when reasoning about action and change. In addition, we acquired a working knowledge of the tools used to do research in this area, temporal logics and preferential entailment. The original approach to solving the frame problem via abnormality theories and nonmonotonic reasoning appeared, on the surface, to be a very good idea. Unfortunately, on closer analysis, it turned out that even for quite straightforward action scenario such as the Yale Shooting Problem, the approach was inadequate. McDermott and Hanks brought the problem into focus with a paper that initiated a great deal of research into reasoning about action and change.

In an attempt to find a satisfactory solution to the Yale Shooting Problem, many researchers presented new approaches to the solution of the Frame problems(s). Each of the solutions can be coarsely classified into three groups:

1. Chronological Minimization – One prefers change to occur as late as possible in time.
2. Causal Minimization – One prefers change that is caused over change without cause.
3. Explanation Closure – One adds axioms to a theory that may be viewed as stating that the only change allowed is that explicitly represented in action rules.

The first two classes generally involve construction of suitable preference criteria for non-monotonic temporal logics. The last class is a "monotonic" approach of sorts where informal nonmonotonic reasoning is used to figure out what additional axioms should be added to a theory which is then used for monotonic reasoning.

In the following two lectures, we will consider a number of different approaches to reasoning about action and change from the various classes. Most of the logics and techniques should be familiar to you from previous lectures.

3 Yale Shooting Problem

McDermott proposes the following axiomatization:

$$Holds(alive, S0) \tag{1}$$

$$\forall s.Holds(loaded, result(load, s)) \tag{2}$$

$$\forall s.Holds(loaded, s) \supset Ab(alive, shoot, s) \wedge Holds(dead, result(shoot, s)) \tag{3}$$

$$\forall f, e, s.Holds(f, s) \wedge \neg Ab(f, e, s) \supset Holds(f, result(e, s)) \tag{4}$$

Keeping in line with our previous representation of abnormality theories, we will use the logically equivalent representation below:

Initial Situation

$$Holds(alive, S0) \tag{5}$$

Laws of Motion

$$\forall s.Holds(loaded, result(load, s)) \tag{6}$$

$$\forall s.Holds(loaded, s) \supset Holds(dead, result(shoot, s)) \tag{7}$$

Frame Axioms

$$\forall f, e, s.Holds(f, s) \wedge \neg Ab(f, e, s) \supset Holds(f, result(e, s)) \tag{8}$$

$$\forall s.Holds(loaded, s) \supset Ab(alive, shoot, s) \tag{9}$$

State Constraints

$$S1 = result(load, S0) \tag{10}$$
$$S2 = result(wait, S1) \tag{11}$$
$$S3 = result(shoot, S2) \tag{12}$$
$$\tag{13}$$

where

$$S3 = result(shoot, result(wait, result(load, S0))) \tag{14}$$
$$\tag{15}$$

The latter axiomatization has simply replaced axiom (3) with axioms (7) and (9) whose conjunction is equivalent to axiom (3). In addition, we introduce additional definitions for situation constants.

We are interested in the *temporal projection problem*. Given a description of the initial situation, the occurrence of some actions, and a description of action laws, what facts are true once the actions have been performed?

Given the current axiomatization, common sense would support the fact that given that the turkey is alive in the initial situation and the action sequence of *load*,*wait*, and $fire$ is performed, the turkey is dead in the resulting situation ($S3$). Using the circumscription policy of minimizing Ab and letting $Holds$ vary, will we be able to infer the intuitive results?

By (5), we know that $Holds(alive, S0)$. By (6) and (10) we know that $Holds(loaded, S1)$. Since there is no evidence to the contrary that $\neg Ab(alive, load, S0)$, we will assume this is the case (C1). (C1) together with (5) and (8) lets us deduce that $Holds(alive, S1)$ (C2). Since there is no evidence to the contrary that the *wait* action will introduce any abnormalities in situation $S1$, we can conjecture that both $\neg Ab(alive, wait, S1)$ (C3) and $\neg Ab(loaded, wait, S1)$ (C4). (C3) and (C4) together with (8) lets us deduce that both $Holds(alive, S2)$ (C5) and $Holds(loaded, S2)$ (C6). (C6) and (7) lets us deduce $Holds(dead, S3)$. (C6) and (9) lets us deduce $Ab(alive, shoot, S2)$ (C7). Since there is no evidence to the contrary that $\neg Ab(loaded, shoot, S2)$ (C8), we can conjecture (C8). (C8) together with (8) and (C6) lets us deduce $Holds(loaded, S3)$.

The model constructed appears to support our intuitions about what should and should not hold as a result of performing the *load*,*wait* and *shoot* actions. In addition, the extension of Ab is minimal, since there is only one tuple $\langle alive, shoot, S2\rangle$ in it and the axiomatization does not allow Ab to have the empty extension. In any model $Holds(loaded, S2)$ is either true or false. If it is true then by axiom (9), $(alive, shoot, S2)$ must be the case. If it is false then since $Holds(loaded, S1)$, $Ab(loaded, wait, S1)$ is the case.

Are these all the Ab-minimal models? Let's try to construct another.

Suppose $Holds(alive, S3)$. Since there is no evidence to the contrary that $\neg Ab(alive, shoot, S2)$ (C1), we can conjecture that this is the case. Taking the contraposition of axiom (9), we can deduce $\neg Holds(loaded, S2)$ (C2). Since $Holds(loaded, S1)$ (C3) may be deduced using axiom (6) and (10), by (C2) and (8) we can deduce $Ab(loaded, wait, S1)$ (C4). In addition, there is no evidence to the contrary that *alive* persists in $S1$ and $S2$, so we can conjecture that $\neg Ab(alive, load, S0)$ and $\neg Ab(alive, wait, S1)$.

Somewhat surprisingly, the model constructed is also Ab-minimal, but does not support our intuitions about what should be concluded from the axiomatization. At best, we can conclude the following:

$$Holds(alive, S3) \vee Holds(dead, S3). \tag{16}$$

The problem of *weak disjunctions* was actually pointed out and known to McCarthy and Lifschitz in earlier work with circumscription, yet the problem also arises in other nonmonotonic formalisms.

4 Post Yale Shooting Problem

Immediately after the appearance of the Yale Shooting Problem, in fact at the same conference in which it was presented, a number of researchers offered *solutions* to the YSP and a debate ensued as to whether these were reasonable solutions or not. Other scenarios were also discovered which played the role of the YSP for these new proposals. This went on for some time and is still going on, but to a lesser extent. The current view emerging from this phase in the history of nonmonotonic reasoning as applied to the domain of action and change is that there is most probably not a general domain independent solution to the frame and related problems, based on a general commonsense logic. But this is another story and we leave it for a future lecture.

In any event, much was learned about the intricacies of reasoning about action and change during this phase and the examples studied offer an insight into many of the problems that would have to be dealt with even for domain dependent solutions to the frame problem(s). This being the case, we will continue by studying a number of proposals for adequately dealing with the YSP and related problems. A coarse division among the proposals can be made in terms of the following classes:

1. Chronological Minimization
2. Causal Minimization
3. Explanation Closure.

5 Chronological Minimization

The previous section provides a hint as to how one might reasonably deal with the YSP. Observe that when constructing the two preferred model classes that the intended models were constructed by moving forward in time and in effect minimizing the Ab predicate in chronological order. The idea is to prefer those models that delay abnormal change as far into the future as possible. In this context, we would prefer those models where $Ab(alive, shoot, S2)$ is true to those where $Ab(loaded, wait, S1)$ is true because the abnormality occurs later in time.

Minimization of the Ab predicate in predicate circumscription is claimed to be inadequate because the minimization is independent of any characteristics of the domain being reasoned about, such as temporal progression. Predicate circumscription prefers those models where the extension of Ab is set-theoretically smaller then the non-preferred models. This minimization has a global flavor to it independent of temporal flow forward in time. Consequently both extensions of Ab

$$Ab = \{\langle alive, shoot, S2\rangle \ldots\}$$

and

$$Ab = \{\langle loaded, wait, S1\rangle \ldots\}$$

in the previous model represent preferred extensions because they can not be compared set theoretically.

The obvious conceptual, but not necessarily heuristic, solution is to define a model preference criteria that captures the idea of minimizing the extension of Ab at each point in time in a forward direction. If we attempt to do this within the framework of abnormality theories and the situation calculus, we will need a more fine-grained type of circumscription than predicate circumscription which compares whole extensions of a predicate. Lifschitz proposed just such a solution and at the same time introduced a new form of circumscription called *pointwise circumscription.*

6 Lifschitz's Pointwise Circumscription Solution

Before considering pointwise circumscription it is useful to review the ways in which the meaning of a predicate constant can be defined in the context of FOL. The most common way to do this is to associate a set of appropriate tuples as the meaning of a predicate constant. In the case of a ternary predicate constant such as Ab with arguments of sort fluent, action and situation, the interpretation of Ab will be a subset of the space of all possible tuples of the proper sort. Another way to do this is to define the meaning of a predicate constant as a characteristic function whose domain is the appropriate tuple space and whose range is true or false. If the tuple belongs to the extension then true is returned, if not then false is returned.

Pointwise circumscription capitalizes on the interpretation of a predicate constant as a characteristic function. The advantage of this representation is that one can now minimize at points in an extension rather than the whole extension of a predicate. In addition, one can also vary or fix other points in an extension while minimizing at a point in the extension. In fact, one need not limit oneself to operating with just one predicate. While minimizing at a point in one predicate, one can specify what part of extensions of other predicates should be fixed or varied within certain constraints. This is of course the analogue to minimizing, varying and fixing various predicate extensions, but in this case we are permitted a more fine-grained minimization policy.

How would this be useful for the present example? Observe that Ab tuples consist of a fluent, action and situation. The idea is to place on ordering on situations and then specify a minimization policy that when minimizing Ab at a point (tuple), fixes those points temporally earlier in the ordering than the situation argument of the current point and varies those points (tuples) whose situation argument is temporally later than the situation argument of the current point being minimized. In addition $Holds$ can vary at all points in its extension.

To make this work, we must make explicit what we mean by "temporally earlier". We do this by adding the following axiom to our premise set,

$$\begin{aligned}\forall s, s'.s < s' \equiv \quad & (s = S0 \wedge s' = S1) \vee \\ & (s = S0 \wedge s' = S2) \vee \\ & (s = S0 \wedge s' = S3) \vee \\ & (s = S1 \wedge s' = S2) \vee \\ & (s = S1 \wedge s' = S3) \vee \\ & (s = S2 \wedge s' = S3)\end{aligned} \tag{17}$$

In addition, we add an axiom defining an ordering on the fluent, action, situation tuples,

$$\forall f, e, s, f', e', s'.\langle f, e, s\rangle <_p \langle f', e', s'\rangle \equiv s < s'. \tag{18}$$

Lifschitz then proposes the following circumscription axiom
$CIRC_{PW}(\Gamma; Ab/ <_P, Holds) \equiv$

$$\forall f, e, s, Ab', Holds'.\neg[Ab(f, e, s) \tag{19}$$
$$\wedge \forall f', e', s'.\langle f', e', s'\rangle <_p \langle f, e, s\rangle$$
$$\supset (Ab'(f', e', s') \equiv Ab(f', e', s')) \tag{20}$$
$$\wedge \Gamma(\lambda f', e', s'.(Ab'(f', e', s') \wedge \langle f, e, s\rangle \neq \langle f', e', s'\rangle), Holds')] \tag{21}$$

The circumscription axiom states that for any substitutions Ab' and $Holds'$ of Ab and $Holds$ and point $\langle f, e, s\rangle$, it can not be the case that

1. Ab is true at $\langle f, e, s\rangle$

   and that

2. Ab and Ab' agree on all points temporally earlier than s

   and that

3. Ab' satisfies the theory Γ at all points other than $\langle f, e, s\rangle$.

6.1 Using Pointwise Circumscription

Although conceptually, pointwise circumscription makes a lot of sense, there is the problem of actually using it to derive intended conclusions. Note that the circumscription axiom is a second-order formula where one quantifies over Ab' and $Holds'$. To use the axiom, we must find an appropriate substitution for Ab' and $Holds'$. This is easier said than done and is one of the criticisms often voiced as regards circumscription in general. In some sense, we must already have a good idea of what our intended conclusions are in order to make the proper substitutions that verify that the conclusions we want can in fact be derived. Although the critics have a point, there are many cases where different forms of circumscription can be reduced to workable forms if the original premise sets to which they are applied meet certain constraints. We will discuss this issue in future lectures. For the time being though, it is worthwhile to go through an example in detail showing how pointwise circumscription could be used.

Step1: Choose substitutions for Ab' and $Holds'$.

We use some wishful thinking and choose substitutions that correspond to the preferred model.

$$\begin{aligned} Ab_c \equiv \lambda f,e,s.(\langle f,e,s\rangle \in \{&\langle alive, shoot, result(shoot, S1)\rangle \\ &\langle alive, shoot, S2\rangle \\ &\langle alive, shoot, S3\rangle \\ &\langle alive, shoot, result(shoot, S3)\rangle\} \end{aligned} \tag{22}$$

$$\tag{23}$$

$$\begin{aligned} Holds_c \equiv \lambda f,s.(\langle f,s\rangle \in \{&\langle alive, S0\rangle \\ &\langle alive, S1\rangle, \langle loaded, S1\rangle \\ &\langle alive, S2\rangle, \langle loaded, S2\rangle \\ &\langle dead, S3\rangle, \langle loaded, S3\rangle \\ &\langle dead, result(shoot, S1)\rangle \\ &\langle dead, result(shoot, S3)\rangle\} \end{aligned} \tag{24}$$

Given the substitutions, we would like to be able to conclude that

$$Ab(loaded, wait, S1) \tag{25}$$

is not true. Let's instantiate the circumscription axiom with this tuple and our substitutions,

$$\neg[Ab(loaded, wait, S1) \tag{26}$$

$$\wedge\forall f',e',s'.\langle f',e',s'\rangle <_p \langle loaded, wait, S1\rangle$$

$$\supset(Ab_c(f',e',s') \equiv Ab(f',e',s')) \tag{27}$$

$$\wedge\Gamma(\lambda f',e',s'.(Ab_c(f',e',s')\wedge\langle loaded, wait, S1\rangle \neq \langle f',e',s'\rangle), Holds')] \tag{28}$$

Let's call the three conjuncts C1', C2' and C3'. If we can show that C2' and C3' are true with the help of Γ then by the circumscription axiom it must follow that $\neg Ab(loaded, wait, S1)$.

In fact, this can be done. We save the details for the lecture with practical exercises.

6.2 Problems

Lifschitz's criteria for preferring models appears to do the job for the Yale Shooting Problem and other problems of Temporal Projection, although it is difficult to say whether there are other counterexamples to this claim at this point. (In fact, there are, as we shall see!). What can be said is that the circumscription axiom produces exactly those theorems true in the chronologically minimal models. The question is whether the criteria is general enough to deal with all types of projection scenarios. The major problem is using the circumscription axiom to generate useful inferences. Our example shows the strategy and difficulties involved.

1. We know what intended conclusions we want from our theory.
2. We use this knowledge to construct an instance of the circumscription axiom.
3. We use this instance together with the other axioms in the theory to see if we can derive the intended conclusions we have in mind.

In effect there is no inferential mechanism for deriving intended conclusions that we do not know about, which is really the point of using logic as an inference mechanism in the first place. What is more important is the actual criteria for preferring models. The criterion of chronological minimization makes sense for temporal projection problems. It would be nice if one could side step the complex inference mechanism associated with circumscription by constructing algorithms that automatically generate the theorems true in preferred models associated with a particular preference criterion.

Shoham has such a proposal, but before considering it, we will look at one other solution to the YSP proposed by Kautz, that is similar to Lifschitz's but does not use the situation calculus. We will also present a scenario in the Kautz context that would cause similar problems in Lifschitz's formalism.

7 Kautz's Logic of Persistence

Kautz proposed a Logic of Persistence to deal with the Yale Shooting Problem which is very similar to Lifschitz's pointwise solution, but is presented using a reified temporal logic with emphasis on the preferential semantics definition for preferring models rather than it's syntactic characterization, although he describes both. The idea is to introduce a predicate constant $Clip$ with two arguments, a fact f and a timepoint t. The intended meaning of $Clip$ is that it holds of a fact f and time t if the fact becomes false at that time. In other words, asserting $Clip(t, f)$ represents a termination or change in value from true to false of a fact. In other words, it no longer persists. In moving forward in time from one timepoint to the next, the value of a fact either persists or is clipped. This information is represented in the form of a persistence axiom

$$\forall t, f. Holds(t, f) \supset Holds(t+1, f) \oplus Clip(t+1, f) \tag{29}$$

which states that if a fact f holds at time t, $Holds(t, f)$, then either it persists, $Holds(t+1, f)$, or its persistence is clipped, $Clip(t+1, f)$. The *oplus* operator is defined as an exclusive-or operator.

Simply circumscribing $Clip$ while varying $Holds$ would still result in two classes of preferred models analogous to those discussed above. In order to avoid the unintended class, the criteria for preferring one model over the other would be to minimize $Clip$ chronologically, preferring models where the $Clip$ instances are as late in time as possible. This of course is quite similar to Lifschitz's approach where Ab, representing abnormal change, is minimized chronologically. Informally, if $M1$ and $M2$ are models then $M1$ is as good as, or better than $M2$, if every clipping instance in $M1$ is matched by an identical clipping instance in $M2$, or by an earlier clipping instance in $M2$ (not necessarily identical to the current clipping instance), which does not also appear in $M1$. Before presenting the formal definition, the axiomatization for the YSP using Kautz's logic is shown below.

Initial Situation

$$Holds(1, loaded) \tag{30}$$

$$Holds(1, alive) \tag{31}$$

Laws of Motion

$$\begin{aligned}&\forall t. Holds(t, fire) \wedge Holds(t, loaded) \supset \\ &\quad \neg Holds(t+1, loaded) \wedge \neg(t+1, alive)\end{aligned} \tag{32}$$

$$Holds(3, fire) \tag{33}$$

Frame Axioms

$$\forall t, f. Holds(t, f) \supset Holds(t+1, f) \oplus Clip(t+1, f) \tag{34}$$

State Constraints

$$\alpha \oplus \beta \equiv_{def} (\alpha \wedge \neg beta) \vee (\neg\alpha \wedge beta) \tag{35}$$

$$loaded \neq alive \wedge loaded \neq fire \wedge alive \neq fire \tag{36}$$

7.0.1 Semantic Definition of Preferential Entailment

Let $M1$ and $M2$ be models and P a predicate of arity n. We write $M1[P]$ to denote the extension of P in $M1$ and $\langle x_1, \ldots, x_n \rangle$ to denote a tuple in $M1[P]$. Besides the predicates $Clip$ and $Holds$, the signature includes the predicate constant $<$ which represents the binary ordering relation on time points. So $t' < t$ denotes that t' is temporally before t. Other domain dependent predicates may also be in the signature. Our minimization policy will minimize $Clip$ chronologically while varying $Holds$ and fixing all other predicates, functions and constants.

Definition 7.1 *$M1$ is preferred to $M2$, written $M1 \sqsubseteq M2$ iff*

1. *$M1$ and $M2$ have the same domain.*
2. *Every constant, function and predicate symbol other than $Clip$ and $Holds$ receives the same interpretation in $M1$ and $M2$.*
3. *The following (meta-thoeretic) statement is true.*

$$\begin{aligned}&\langle f, t \rangle \in M1[Clip] \supset (\langle f, t \rangle \in M2[Clip] \vee \\ &(\exists t', f'. \langle f, t \rangle \in M2[Clip] \wedge \langle f', t' \rangle \notin M1[Clip] \wedge \\ &\qquad \langle t', t \rangle \in M1[<]))\end{aligned} \tag{37}$$

A model $M1$ is said to be strictly preferred *to $M2$, written $M1 \sqsubset M2$, if $M1 \sqsubseteq M2$ and not $M2 \sqsubseteq M1$.*

The chronologically minimal models are those that are strictly preferred. Observe that this a partial ordering so there may be many minimal models. Intuitively, if $M1 \sqsubset M2$ then the extension of $Clip$ is identical in $M1$ and $M2$ up to some time t' and at t' the extension of $Clip$ in $M2$ properly includes the extension of $Clip$ in $M1$, $M1[Clip] \subset M2[Clip]$. Since $Holds$ is allowed to vary, its extension at points before or after t' are irrelevant to the preference ordering.

7.0.2 Syntactic Characterization of Preferential Entailment

The semantic definition above which characterizes the chronologically minimal models is straightforward and provides a clear intuitive description of what is meant by a chronologically minimal model. This does not help us compute inferences. Like Lifschitz, Kautz also provides a syntactic characterization of his definition of preferential entailment by using a circumscription axiom

$$\begin{aligned} \forall Clip', Holds'.[\Gamma(Clip', Holds')\wedge \\ \forall t, f.Clip'(t,f)\supset[Clip(t,f)\vee \\ \exists t', f'.t' < t\wedge Clip(t',f')\wedge\neg Clip'(t',f')]] \\ \supset\forall t, f.Clip(t,f)\equiv Clip'(t,f) \end{aligned} \tag{38}$$

It is useful to compare Kautz's circumscription axiom with Lifschitz's.

7.0.3 Using Kautz's Persistence Circumscription

The example is considered in the paper and will be done during the practical lecture.

7.0.4 Problems

Kautz's logic of persistence certainly provides the intuitive results for the YSP as axiomatized, and the intuitions behind chronological minimization of clipping are clear. Yet, there are certainly a number of other scenarios where one would hope that this approach provides the proper results, but doesn't. Kautz mentions a few of the problems himself. Here is one.

8 Shoham's Logic of Chronological Ignorance

Shoham constructs a nonmonotonic logic called *chronological ignorance*(CI) which combines elements of temporal logic, modal logic and nonmonmotonic logics. He also uses the strategy described in a previous lecture of taking an arbitrary logic and defining a preference ordering on its models. He then supplies an algorithm for computing preferred entailments directly. We will introduce Shoham's logic in two steps:

1. Describe a monotonic modal logic of temporal knowledge.
2. Describing a partial order on its models.

Recall Shoham's previous temporal logic STL, discussed in an earlier lecture. Temporal information was described in terms of logical combinations of atomic formulas of the following type

$$TRUE(t_1, t_2, r(\mu_1, \ldots, \mu_n),$$

where r is a relation symbol and $\mu_1, \ldots, \mu_n$ are atemporal terms. In CI Shoham basically uses the same representation of temporal information but the propositional type is simplified to be a propositional constant rather than allowing arguments

$$TRUE(t_1, t_2, p).$$

In addition to the temporal information, Shoham introduces a modal operator, $\Box$ which may loosely be understood as a *knowledge* operator. So,

$$\Box TRUE(t_1, t_2, p)$$

asserts that p is known to be true during the interval from t_1 to t_2. Shoham also assumes an integer time structure $\langle \mathcal{N}, \leq \rangle$. Models or interpretations would then be similar to *Kripke interpretations* which are pairs $\langle W, M \rangle$ where W is a non-empty universe of (possible) worlds, and M is a meaning function. Possible worlds can be viewed as classical models with a integer time structure which is the same for all worlds. In other words, a set of infinite parallel time lines all sharing the same temporal interpretation but with different facts true at different points in time. Each possible world then represents a possible evolution in time or history of the universe. We provide the formal definitions below

8.1 TK: A Monotonic Logic of Temporal Knowledge

Syntax:

1. **P**: A set of primitive propositions.
2. **TV**: A set of temporal variables.
3. **TC**: The set $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$.
4. **U**= **TC**$\cup$ **TV**: The set of temporal terms.

The set of well-formed formulas is defined inductively as follows:

1. If $u_1 \in \mathbf{U}$ and $u_2 \in \mathbf{U}$ then $u_1 = u_2$ and $u_1 \preceq u_2$ are wffs.
2. If $u_1 \in \mathbf{U}$, $u_2 \in \mathbf{U}$, and $p \in \mathbf{P}$ then $TRUE(u_1, u_2, p)$ is a wff.
3. If α and β are wffs then so are $\alpha \wedge \beta$,$\neg\alpha$, and $\Box\alpha$.
4. If α is a wff and $v \in \mathbf{TV}$, then $\forall v.\alpha$ is also a wff.

In addition there are a number of abbreviations that may be found in the article.

Semantics:

Shoham assumes an S5 modal structure with a fixed interpretation of time across worlds. That fixed interpretation is the integer structure $\langle \mathcal{N}, \leq \rangle$.

Definition 8.1 (Kripke Interpretation) *A* Kripke Interpretation *is a pair $\langle W, M \rangle$ where W is a non-empty universe of possible worlds, and M is a meaning function $M : \mathbf{P} \to 2^{W \times \mathcal{N} \times \mathcal{N}}$. In addition, $\langle w, t, t' \rangle \in M(p)$ iff $\langle w, t', t \rangle \in M(p)$.*

Definition 8.2 (Variable Assignment) *A* variable assignment *is a function $VA : \mathbf{TV} \to \mathcal{N}$.*

If $u \in \mathbf{U}$ then $Val(u) = VA(u)$ if $u \in \mathbf{TV}$ and the standard interpretation *of u if $u \in \mathbf{TC}$.*

Definition 8.3 (Satisfaction Conditions) *A Kripke Interpretation $KI = \langle W, M \rangle$ and a world $w \in W$ satisfy a formula α under the variable assignment VA, written $KI, w \models \alpha[VA]$, under the following conditions:*

- $KI, w \models u_1 = u_2[VA]$ *iff* $VAL(u_1) = VAL(u_2)$.
- $KI, w \models u_1 \preceq u_2[VA]$ *iff* $VAL(u_1) \leq VAL(u_2)$.
- $KI, w \models TRUE(u_1, u_2, p)[VA]$ *iff* $\langle w, VAL(u_1), VAL(u_2) \rangle \in M(p)$.
- $KI, w \models \alpha \wedge \beta[VA]$ *iff* $KI, w \models \alpha[VA]$ *and* $KI, w \models \beta[VA]$

- $KI, w \models \neg\alpha[VA]$ *iff* $KI, w \not\models \alpha[VA]$.
- $KI, w \models (\forall v\alpha)[VA]$ *iff* $KI, w \models \alpha[VA']$ *for all* VA' *that agree with* VA *everywhere except possible on* v.
- $KI, w \models \Box\alpha[VA]$ *iff* $KI, w' \models \alpha[VA]$ *for all* $w \in W$.

Definition 8.4 (Semantic Entailment) *A Kripke interpretation* $KI = \langle W, M \rangle$ *and a world* $w \in W$ *are a* model *for a formula* α, *written* $KI, w \models \alpha$, *if* $KI, w \models \alpha[VA]$ *for any variable assignment* $[VA]$.

A wff is satisfiable *if it has a model, and* valid *if its negation has no model.*

$\alpha_1 entails \alpha_2$, *written* $\alpha_1 \models \alpha_2$ *iff* α_2 *is satisfied by all models of* α_1.

The syntax and model theory of TK will be used for representing and interpreting scenario descriptions. As was the case for other approaches, some type of minimization or preference policy will be required to satisfactorily deal with the frame problems(s). Before discussing the preference ordering on Kripke interpretations that will be needed, we first represent a scenario used by Shoham as a vehicle for discussing the frame and qualification problems.

Boundary Conditions

$$\Box(1, loaded) \tag{39}$$

$$\Box(5, fire) \tag{40}$$

Causal Rules

$$\begin{aligned} \Box(t, loaded) \wedge \Box(t, fire) \wedge \\ \Diamond(t, air) \wedge \Diamond(t, firingpin) \\ \wedge \ldots \Diamond \ldots \text{ other mundane conditions} \\ \supset \Box(t+1, noise), \text{ for all } t \in TC. \end{aligned} \tag{41}$$

Frame Axioms

$$\begin{aligned} \Box(t, loaded) \wedge \Box(t, fire) \wedge \\ \Diamond(t, \neg fire) \wedge \Diamond(t, \neg emptied - manually) \\ \supset \Box(t+1, loaded), \text{ for all } t \in TC. \end{aligned} \tag{42}$$

This is the axiomatization Shoham uses in his article. An alternate axiomatization, more in keeping with the abnormality theories used in the situation calculus would use the following axioms instead.

Modified Causal Rules

$$\begin{aligned} \Box(t, loaded) \wedge \Box(t, fire) \wedge \\ \wedge \Diamond(t, sit - normal) \\ \supset \Box(t+1, noise), \text{ for all } t \in TC. \end{aligned} \tag{43}$$

$$\Box(t, \neg air) \supset \Box(t, \neg sit - normal), \text{ for all } t \in TC. \tag{44}$$

$$\Box(t, \neg firingpin) \supset \Box(t, \neg sit - normal), \text{ for all } t \in TC. \tag{45}$$

$$\ldots \Box \ldots \text{ other mundane conditions} \ldots \supset \Box(t, \neg sit - normal), \text{ for all } t \in TC. \tag{46}$$

<u>**Modified Frame Axioms**</u>

$$\begin{aligned} &\Box(t, loaded) \wedge \Box(t, fire) \wedge \\ &\wedge \Diamond(t, sit - normal_1) \\ &\supset \Box(t+1, loaded), \text{ for all } t \in TC. \end{aligned} \tag{47}$$

$$\Box(t, \neg fire) \supset \Box(t, \neg sit - normal_1), \text{ for all } t \in TC. \tag{48}$$

$$\Box(t, \neg emptied - manually) \supset \Box(t, \neg sit - normal_1), \text{ for all } t \in TC. \tag{49}$$

Note the explicit use of a frame axiom rather than a general frame rule. Although this might provide a solution to the "extended prediction" problem, which supposedly subsumes the frame problem, there is still the unsatisfying requirement of one frame axiom and a number of additional exception axioms for *each* fluent such as *loaded*. To be fair, Shoham is aware of this and the paper you received deals more specifically with the qualification problem, or as he calls it the *initiation problem*. But even in this case, due to the use of schemas (to insure that we reason propositionally), the resulting axiomatization is not as succinct as one might wish. On the other hand, such an axiomatization is advantageous in regard to algorithms for constructing preferred models directly.

8.2 CI: A Nonmonotonic Logic of Temporal Knowledge

CI, the logic of chronological ignorance, is a nonmonotonic version of TK with the same syntax, but also augmented with an ordering relation on Kripke structures.

Definition 8.5 (Base Wff) *A* base wff *is a wff containing no modal operator.*

Let $maxt(t_1, t_2)$ be the max function for time points relative to the ordering $\leq$.

Definition 8.6 (Latest Time Point) *The* latest time point (l.t.p) *of a base formula is the latest time point mentioned in it defined as follows.*

1. *The l.t.p of $TRUE(t_1, t_2, p)$ is $max(t_1, t_2)$.*
2. *The l.t.p of $\alpha \wedge \beta$ is the maxt of the l.t.p's of α and β.*
3. *The l.t.p of $\neg\alpha$ is the l.t.p of α.*
4. *The l.t.p of $\forall v.\alpha$ is the earliest among the l.t.p's of all α' which result from substituting in α a time point symbol for all free occurrences of v, or $-\infty$ if there is no such earliest point.*

Definition 8.7 (Chronologically more ignorant) *A Kripke interpretation M_2 is* chronologically more ignorant *than a Kripke interpretation M_1, written $M_1 \sqsubseteq M_2$, if there exists a time t_0 such that*

1. *for any base sentence α whose l.t.p. $\leq t_0$, if $M_2 \models \Box\alpha$ then also $M_1 \models \Box\alpha$, and*
2. *there exists some base sentence α whose l.t.p. is t_0 such that $M_1 \models \Box\alpha$ but $M_2 \not\models \Box\alpha$.*

Definition 8.8 (Chronologically Maximally Ignorant) *M is said to be a* chronologically maximally ignorant*(c.m.i.) model of α if $M \models_{\sqsubset_{ci}} \alpha$, that is, if $M \models \alpha$ and there is no other M' such that $M' \models \alpha$ and $M \sqsubset_{ci} M'$.*

Definition 8.9 (CI) *The logic of chronological ignorance, CI, is the nonmonotonic logic $TK_{\sqsubset_{ci}}$.*

8.2.1 Format of Axioms

Axioms in scenario descriptions like the YSP are generally restricted to two forms:

1. Boundary conditions represented as knowledge about atomic sentences.

$$\Box(1, loaded)$$

2. Causal rules and Frame axioms represented as implications from both knowledge $\Box$ and ignorance ($\Diamond \equiv \neg\Box\neg$) of past atomic sentences to knowledge of future atomic sentences.

$$\begin{aligned}\Box(t, loaded) \wedge \Box(t, fire) \wedge \\ \Diamond(t, air) \wedge \Diamond(t, firingpin) \\ \wedge \ldots \Diamond \ldots \text{ other mundane conditions} \\ \supset \Box(t+1, noise), \text{ for all } t \in TC.\end{aligned} \tag{50}$$

The advantage of limiting axioms to these formats is that for any theory, the base sentences will be the same in any c.m.i. model for the theory. Lets provide a formal definition for theories constructed from the restricted language. Shoham calls these *causal theories*.

Definition 8.10 (Causal Theory) *A* causal theory *Ψ is a theory in CI, in which all sentences have the form*

$$\Phi \wedge \Theta \supset \Box\alpha$$

where

1. *α is a (positive or negative) atomic base sentence.*
2. *Φ is a conjunction of sentences $\Box\alpha_i$, where α_i is a (positive or negative) atomic base sentence whose l.t.p. $<$ the l.t.p. of alpha.*
3. *Θ is a conjunction of sentences $\Diamond\alpha_i$, where α_i is a (positive or negative) atomic base sentence whose l.t.p. $<$ the l.t.p. of alpha.*
4. *Either or both Φ and Θ may be empty (true). A sentence in which Φ is empty is called a* boundary condition*. Other sentences are called* causal rules.
5. *There is a time point $t_0 \neq -\infty$ preceding the l.t.p. of the r.h.s. in all boundary conditions (that is, if $\Theta \supset \Box\alpha$ is a boundary condition, then $t_0 <$ the l.t.p. of α.*
6. *For any p, t_1, t_2, there do not exist two sentences in Ψ such that one contains $\Diamond(t_1, t_2, p)$ on its l.h.s. and the other contains $\Diamond(t_1, t_2, \neg p)$ on its l.h.s.*

7. *If* $\Phi_1 \wedge \Theta_1 \supset \Box(t_1, t_2, p)$ *and* $\Phi_2 \wedge \Theta_2 \supset \Box(t_1, t_2, \neg p)$ *are two sentences in* Ψ*, then* $\Phi_1 \wedge \Theta_1 \wedge \Phi_2 \wedge \Theta_2$ *is inconsistent.*

Why are causal theories so nice? They have a number of properties that make it easy to construct c.m.i. models.

Theorem 8.1 *If* Ψ *is any causal theory, then*

1. Ψ *has a c.m.i. model,*
2. *If* M_1 *and* M_2 *are both c.m.i. models of* Ψ*, and* α *any base sentence, then* $M_1 \models \Box\alpha$ *iff* $M_2 \models \Box\alpha$*, and*
3. *If* M *is a c.m.i. of* Ψ*, then the set of known base sentences consists of the tautological consequences of the (positive and negative) known atomic base sentences.*

What this theorem states is that for any causal theory Ψ, its preferred models will have the same set of base sentences, the theory will be consistent, and the models may be built by moving forward in time, adding only those sentences that we know about and those implied by what is already known. We can also do this efficiently for each causal theory.

8.2.2 Intuitions behind the logic

As Shoham states, the intuition behind his logic and causal theories is the "what you do not know will not hurt you" principle. Note that $\Diamond\alpha \equiv \neg\Box\neg\alpha$, $\Diamond\alpha$ atoms are used to qualify causal rules, and we minimize $\Box$. The net result is that unless we are explicitly aware that certain conditions hold, we may assume we do not know about them and the causal rules will have there normal effects. Of course, if we later learn to know of a condition, or conditions, that do qualify a causal rule, then these should be explicitly added to out axiomatization. In effect, we only include conditions in our axiomatization that we explicitly know about. We can disregard all other qualifications.

Shoham's logic exploits the partiality of knowledge contexts. Since the logic is S5-based, it is the case that

$$\models \Box(t, p) \vee \neg\Box(t, p)$$

but ,

$$\not\models \Box(t, p) \vee \Box(t, \neg p).$$

This is important to understand because the problem of weak disjunctions in the YSP has to do with the fact that when Ab is minimized, $\neg Ab$ is maximized. So minimizing "truth" does not minimize all propositions simultaneously. On the other hand, minimizing knowledge ($\Box$), does minimize all propositions simultaneously within the knowledge context. If one minimizes knowledge about p, then $\neg\Box(t, p)$ does not imply that $\Box(t, \neg p)$, it may also be the case that $\neg\Box(t, \neg p)$.

In the YSP axiomatization described by McDermott, the following disjunction is implied,

$$Ab(alive, shoot, S2) \vee Ab(loaded, wait, S1).$$

Minimizing Ab globally would mean that either

$$\neg Ab(alive, shoot, S2) \wedge Ab(loaded, wait, S1)$$

or

$$Ab(alive, shoot, S2) \wedge \neg Ab(loaded, wait, S1)$$

would be true in any Ab-preferred model. This results in the generation of two classes of preferred models, those containing the former conjunction are unintended. The solution proposed is to use the situation arguments as "hooks" for prioritizing and the fact that $S1 < S2$, to minimize Ab in chronological order. Using Shoham's approach, where knowledge is minimized rather than truth, there are no problems as far as one minimization influencing another. Minimizing *knowledge* of the truth or falsity of a proposition does not affect knowledge about any other propositions, except indirectly, where knowledge about propositions later in time can be derived as an indirect result of applying the causal rules together with our conjectures of ignorance.

8.2.3 The YSP Example

Let's take a look an an axiomatization of the YSP and describe what the c.m.i. models would include.

Boundary Conditions

$$\Box(0, loaded) \tag{51}$$

$$\Box(0, alive) \tag{52}$$

$$\Box(2, fire) \tag{53}$$

Causal Rules

$$\begin{aligned}\Box(t, loaded) \wedge \Box(t, fire) \wedge \\ \supset \Box(t+1, \neg alive), \text{ for all } t \in TC.\end{aligned} \tag{54}$$

Frame Axioms

$$\begin{aligned}\Box(t, loaded) \wedge \Diamond(t, \neg fire) \wedge \Diamond(t, \neg emptied - manually) \\ \supset \Box(t+1, loaded), \text{ for all } t \in TC.\end{aligned} \tag{55}$$

It is easy to see that, if starting at time point 0 and reasoning forwards, all c.m.i models of the theory will satisfy $\Box(3, \neg alive)$. We can show this by constructing the sentences known at each time point.

- **Tp 0:** Boundary conditions: $\Box(0, loaded)$, $\Box(0, alive)$.
- **Tp 1:** Instantiate axiom (55)

$$\begin{aligned}\Box(0, loaded) \wedge \Diamond(0, \neg fire) \wedge \Diamond(0, \neg emptied - manually) \\ \supset \Box(1, loaded),\end{aligned} \tag{56}$$

or equivalently,

$$\begin{aligned}\Box(0, loaded) \wedge \neg\Box(0, fire) \wedge \neg\Box(0, emptied - manually) \\ \supset \Box(1, loaded).\end{aligned} \tag{57}$$

By minimizing $\Box$ at Tp 1, we can derive $\neg\Box(0, fire)$, $\neg\Box(0, emptied - manually)$. Consequently, we can derive $\Box(1, loaded)$ using modus ponens.

Proof: Suppose $M \models \Box(1, loaded)$ and $M' \not\models \Box(1, loaded)$. By axiom (55), either $M' \models \Box(0, fire)$ or $M' \models \Box(0, emptied - manually)$. Based on the c.m.i. definition, it must be the case that $M' \sqsubseteq M$ which implies that M is a c.m.i. model and $\Box(1, loaded)$ is true in it.

- **Tp 3:** A similar argument can be used to derive $\Box(2, loaded)$ (d1). (d1) together with axioms (53) and (54) permit us to derive $\Box(3, \neg alive)$ in all c.m.i. models as desired.

8.2.4 Problems

Using modal logics of the sort Shoham proposes, sometimes hide subtle anomalies which are not only hard to find, but also result in characteristics which run counter-intuitive to the phenomena being modeled. For example, Shoham points out a particular property of his causal theories that violates the supposed universality of his causal rules, which are meant to describe a physics of sorts, and consequently should behave reasonably in all possible worlds. This is desirable, because one of those possible worlds may in fact be the actual world and we would hope that the actual world provides a physics that reflects our common sense ideas of what should apply. Recall the following somewhat simplified causal rule:

$$\begin{aligned} \Box(t, loaded) \wedge \Box(t, fire) \wedge \\ \Diamond(t, air) \\ \supset \Box(t+1, noise), \text{ for all } t \in TC. \end{aligned} \tag{58}$$

or the equivalent,

$$\begin{aligned} \Box(t, loaded) \wedge \Box(t, fire) \wedge \\ \neg\Box(t, \neg air) \\ \supset \Box(t+1, noise), \text{ for all } t \in TC. \end{aligned} \tag{59}$$

The idea is that unless a loaded gun is fired in a vacuum ($\Box(t, \neg air)$), a noise should follow. When knowledge ($\Box$) is minimized, $\neg\Box(t, \neg air)$ becomes true. Under the assumption that both $\Box(t, loaded)$ and $\Box(t, fire)$ are true, we can also derive $\Box(t+1, noise)$. So, given a Kripke structure that is preferred, it is the case that $TRUE(t+1, noise)$ holds in all possible worlds in the structure where $\Box(t, loaded)$ and $\Box(t, fire)$ are true. So far, so good. But what about, $TRUE(t, air)$? Since we know that $\neg\Box(t, \neg air)$ is true in the structure, there can be possible worlds where $TRUE(t, \neg air)$ holds and others where $TRUE(t, air)$ holds. But this would mean it is possible that in one of our preferred Kripke structures there is a possible world where both $TRUE(t, \neg air)$ and $TRUE(t+1, noise)$ both hold. This of course runs counter to the implicit assumption of the universality of the physics or causal rules. That particular possible world might just be the actual world and there would be noise in a vacuum.

Realizing this, Shoham introduces an additional *soundness condition* whose purpose is to rule out possible worlds in preferred Kripke structures of the type we just described.

Definition 8.11 (Soundness Condition) *The* soundness conditions *of a causal theory* Ψ *are the set of sentences*

$$\Diamond(t_1, t_2, p) \supset TRUE(t_1, t_2, p)$$

such that $\Diamond(t_1, t_2, p)$ *appears on the l.h.s. of some sentence in* Ψ.

The soundness conditions are implicitly assumed to be part of the causal theory itself, but are left out for economy of representation. So, given that the causal rule above is part of our causal theory,

$$\Diamond(t, air) \supset TRUE(t, air)$$

would be a soundness condition of the theory. In any possible world in a Kripke structure where $\Diamond(t, air)$ ($\neg\Box(t, \neg air)$), the soundness condition will constrain the world to include $TRUE(t, air)$.

This soundness condition appears to have been added as an afterthought and appears somewhat arbitrary. Perhaps a good way to explain it after the fact is as follows. The soundness condition

$$\neg\Box(t, \neg air) \supset TRUE(t, air),$$

can be paraphrased as representing the fact that

> "Unless I know that there is a vacuum, I will assume this is not the case and reason *as if* it were in fact true that there is air.

In other words, one uses a default as if it were true. Perlis in another article argues that this is the way a default should in fact be interpreted: as a two step process in which one first jumps to a conclusion or default, and secondly uses it as if it were in fact a legitimate assertion.

There are a number of other anomalies associated with Shoham's approach, but we will save these for a future lecture.