PROCEEDINGS OF THE INTERNATIONAL CONFERENCE
ON FIFTH GENERATION COMPUTER SYSTEMS 1992,
edited by ICOT. © ICOT, 1992

A Mechanism for Reasoning about Time and Belief

Hideki Isozaki
NTT Basic Research Laboratories
3–9–11, Midoricho, Musashino-shi
Tokyo 180, Japan

Yoav Shoham
Computer Science Department
Stanford University
Stanford, CA 94305, U.S.A.

Abstract

Several computational frameworks have been proposed to maintain information about the evolving world, which embody a *default persistence* mechanism; examples include *time maps* and the *event calculus*. In multi-agent environments, time and belief both play essential roles. Belief interacts with time in two ways: there is the time *at* which something is believed, and the time *about* which it is believed.

We augment the default mechanisms proposed for the purely temporal case so as to maintain information not only about the objective world but also about the evolution of beliefs. In the simplest case, this yields a two-dimensional map of time, with persistence along each dimension.

Since beliefs themselves may refer to other beliefs, we have to think of a statement referring to an agent's temporal belief about another agent's temporal belief (*a nested temporal belief statement*). It poses both semantical and algorithmic problems. In this paper, we concentrate on the algorithmic aspect of the problems. The general case involves multi-dimensional maps of time called *Temporal Belief Maps*.

1 Introduction: *Time Maps* and *Temporal Belief Maps*

In multi-agent environments, time and belief both play essential roles. Belief interacts with time in two ways: there is the time *at* which something is believed, and the time *about* which it is believed. As in the atemporal treatment of belief, beliefs themselves may refer to beliefs (of other agents, or even the same one). For example, in the framework of Agent Oriented Programming [Shoham 1990], at any time the mental state of an agent contains information about the mental states of other agents at various times.

A statement referring to an agent's temporal belief about another agent's temporal belief will be called a *nested temporal belief statement*. An example of it is the sentence "On Wednesday John believed that on the previous Monday Jane believed that on the following Saturday they would clean the house." Nested temporal beliefs pose a number of interesting problems, both semantical and algorithmic. In this paper we concentrate on the latter kind; we propose a computational mechanism called a *Temporal Belief Map*, which functions as a data base of nested temporal beliefs.

Consider a formal language for expressing nested temporal beliefs. A standard construction would extend classical logic with a modal operator $\mathtt{B}_a^t\varphi$ for each agent designator a and time point symbol t, meaning intuitively that at time t the agent a believes φ. To ensure that the modal operator respects the properties of belief (or, more exactly, its crude approximation that has been employed in computer science and AI), various restrictions on this operator have been suggested, and then extensively explored, debated and modified[Hintikka 1962, Griffiths 1967, Konolige 1986]. These include properties such as $\mathtt{B}_a^t(\varphi \supset \psi) \wedge \mathtt{B}_a^t\varphi \supset \mathtt{B}_a^t\psi$ (the 'K' axiom), $\mathtt{B}_a^t\varphi \supset \neg\mathtt{B}_a^t\neg\varphi$ (the 'D' axiom), $\mathtt{B}_a^t\varphi \supset \mathtt{B}_a^t\mathtt{B}_a^t\varphi$ and $\neg\mathtt{B}_a^t\varphi \supset \mathtt{B}_a^t\neg\mathtt{B}_a^t\varphi$ (the '4' and '5' axioms)[Chellas 1980], and others. In addition, although these have been less well studied, further constraint may be imposed on the *change* in belief over time.

We will briefly return to these properties in the next section, but they are not the focus of this paper. Instead, we concentrate on algorithmic issues. Consider first the purely temporal case, without an explicit notion of belief. In principle, capturing the truth of facts over time should pose no problem; we can use standard data base techniques to capture the fact true at a single point in time, and repeat it for all point. In practice, though, it is impossible, and we will need to use some shortcuts. The representational aspect of the problem appears in the form of the well-known *frame problem* [McCarthy and Hayes 1969]: when you buy a red bicycle, how you conclude that a year later it will still be red, regardless of what happens in the meanwhile — the bike is ridden, the tire is fixed, elections are held — unless it is painted. An axiom stating explicitly that the color does not *change* after each action is called a *frame axiom*; the problem is to capture the persistence of facts without including the numerous possible frame axioms.

The frame problem and related problems have been investigated in detail from the logical point of view (cf. [Shoham 1992]), and most solutions proposed have made use of nonmonotonic logic. Adding belief yields a qualitative increase in difficulty, since beliefs (and lack thereof) tend to persist as well: once you learn something, you will keep it in mind until you forget it or learn incompatible facts. The formal details of the persistence of mental state have not yet been studied as deeply; an initial treatment of it appears in [Lin and Shoham 1992].

As was said, we are interested in the algoirthmic aspects of the problem. Computational complexity of knowledge and belief without time was discussed by [Halpern and Moses 1985]. In the purely temporal case, the question is how to efficiently implement the following persistence principle (throughout this article we will assume discrete time, but the discussion can be adapted to the continuous case as well; we also assume propositional facts, with no variables):

> p^{t+1} holds iff either an event which causes p occurred at time t, or else p^t holds and no event which causes $\neg p$ occurred at time t.

Straightforward embodiment of this rule in backward chaining is too inefficient. In order to determine the truth value of p^t, you do not want to have to check p^{t-1}, p^{t-2}, and so on until you discover that $p^{t-213857}$ is true.

Both *time maps* [McDermott 1982, Dean and McDermott 1987] and *event calculus* [Kowalski and Sergot 1986] provide better alternatives. In particular, time maps rely on keeping track of only the points at which the truth value of the proposition changes, which are sufficient to

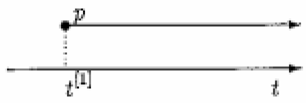

Figure 1: A simple persistence

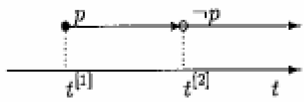

Figure 2: Clipping a persistence

determine the truth value of all other points. Each event gives rise to a *default persistence*, which ends at the first future point about which a contradictory fact is believed. For example, if an event which causes p occurs at time $t^{[1]}$ (the superscript in $^{[\ ]}$ identifies a given time point), and no other information about p is yet present in the time map, then the two points $t^{[1]}$ and ∞ are associated with p, with a default persistence of p from the first to the second. This may be depicted graphically by Figure 1.

If it is subsequently added that at time $t^{[2]}(> t^{[1]})$ an event happened that causes $\neg p$, $t^{[2]}$ is associated in addition with p; a default persistence of $\neg p$ is assumed between $t^{[2]}$ and ∞, and the persistence of p starting at $t^{[1]}$ is "clipped" at $t^{[2]}$ (Figure 2).

This is a crude description of the operation of time maps, but it suffices to explain the transition to *temporal belief maps* (TBM's), which incorporate an explicit notion of belief.

(Note that we have discussed only persistence into the future. Most of the literature in AI does that, and we too will in this paper. However, persistence into the past can make as much sense, especially when one adds an explicit notion of belief. For example, if you find a book on a desk, you will believe that the book was on the desk a few minutes ago. Most researchers manage to avoid this issue by limiting the form of temporal information. In particular, both time maps and the event calculus embody a certain *causality principle*: the only way new temporal information is added is by a preceding event which *causes* it. Since an explicit cause is known, there is no reason to posit backward persistence, past the cause. For example, we cannot represent the simple fact that the book *was* on the table; we must represent a specific event or action that resulted in that state (such as placing the book there). The closest one gets to backward persistence is through abductive reasoning, "what would have to be the case previously in order for this fact to hold," positing previous events. In applications such as planning[Allen et al. 1991], this is a reasonable assumption, as in those one is constructing a map of the future based on specific planned events. However, if one is trying to use the mechanism to piece together a map of time on the basis of spotty data, this may prove inappropriate. For example in a framework such as Agent Oriented Programming [Shoham 1990], a major source of new temporal information are INFORM messages from other agents. As a result of these messages, the agent may possess a rich sample of what is true and false over time, but no causal knowledge of the precipitating events. Nevertheless, we will ignore backward persistence in most of the paper. Unless we explicitly state otherwise, we will use the term *persistence* to mean forward persistence.)

Suppose we now wish to represent the evolution of an agent's beliefs. Let us first introduce the notion of *learning*, which will play a role that is analogous to that of an event in time maps. Given this notion, beliefs too will be subject to a persistence rule:

> "The agent believes a fact at time $t+1$ iff he learned it at time t, or else at time t he believed the fact and did not at that time learn that it became false."

(This rule embodies the assumption that agents have perfect memory.) If, in addition, the "fact" itself is temporal, we end up with persistence along two orthogonal dimensions: the time of belief and the time of the property. This is the simple case of a 2-dimensional TBM.

The extension to higher-dimensional TBM's is natural. Such TBM's are obtained by nested belief statements, such as as "John believes today that yesterday he did not believe ..." and "John believes today that tomorrow Mary will believe ..."); both of these example statements induce a 3-dimensional TBM. It turns out that resolving contradictions in a multi-dimensional TBM is somewhat more subtle than in standard time maps, as the following sections will describe.

Here then is the problem we will address. Let us use the notation $\mathrm{L}_a^t\varphi$ to mean that agent a learned φ at t (actually formalizing this notion is tricky, but that is not the concern of this paper; we use the notation merely as shorthand for the English sentence). The input to our problem is assumed to be a collection of data points of the form $\mathrm{L}_{a_1}^{t_1^{[i]}}\mathrm{L}_{a_2}^{t_2^{[i]}}\cdots\mathrm{L}_{a_{n-1}}^{t_{n-1}^{[i]}}p_i^{t_n^{[i]}}$ and $\mathrm{L}_{a_1}^{t_1^{[j]}}\mathrm{L}_{a_2}^{t_2^{[j]}}\cdots\mathrm{L}_{a_{n-1}}^{t_{n-1}^{[j]}}\neg p_j^{t_n^{[j]}}$. In other words, the sequences of agent indices are identical in all the input data, but the time indices are unconstrained (we will see in section 5 why assuming a fixed sequence of agent indices is not limiting). We also assume that the data is *consistent*, that is, it does not contain both $\mathrm{L}_{a_1}^{t_1^{[k]}}\cdots\mathrm{L}_{a_{n-1}}^{t_{n-1}^{[k]}}p_k^{t_n^{[k]}}$ and $\mathrm{L}_{a_1}^{t_1^{[k]}}\cdots\mathrm{L}_{a_{n-1}}^{t_{n-1}^{[k]}}\neg p_k^{t_n^{[k]}}$ for any k. The problem is to define the rules of persistence in this n-dimensional space, that is, to define for any $(t_1, t_2, \ldots, t_n)$ in the space and each fact p, which (if either) of $\mathrm{B}_{a_1}^{t_1}\mathrm{B}_{a_2}^{t_2}\cdots\mathrm{B}_{a_n}^{t_{n-1}}p^{t_n}$ and $\mathrm{B}_{a_1}^{t_1}\mathrm{B}_{a_2}^{t_2}\cdots\mathrm{B}_{a_n}^{t_{n-1}}\neg p^{t_n}$ are supported by the data. (In all of the above, both the agent indices and the time indices may contain repetitions.) Furthermore, we will want our definition to support an efficient mechanism for answering such a query about any point in the space.

Note that both the input form and query form are quite constrained. For example, the input form precludes facts such as "John learned that Mary did not believe φ," $(\mathrm{L}_{\mathrm{John}}\neg\mathrm{B}_{\mathrm{Mary}}\varphi)$ without making the stronger statement "John learned that Mary learned $\neg\varphi$." $(\mathrm{L}_{\mathrm{John}}\mathrm{L}_{\mathrm{Mary}}\neg\varphi)$

Similarly, a query "Does John believe that Mary does not believe φ?" $(\mathrm{B}_{\mathrm{John}}\neg\mathrm{B}_{\mathrm{Mary}}\varphi?)$ are disallowed, only the stronger query about Mary's believing the negated fact $(\mathrm{B}_{\mathrm{John}}\mathrm{B}_{\mathrm{Mary}}\neg\varphi?)$. A positive answer to the second $(\mathrm{B}_{\mathrm{John}}\mathrm{B}_{\mathrm{Mary}}\neg\varphi)$ would entail a positive one to the first $(\mathrm{B}_{\mathrm{John}}\neg\mathrm{B}_{\mathrm{Mary}}\varphi)$, but a negative answer $(\neg\mathrm{B}_{\mathrm{John}}\mathrm{B}_{\mathrm{Mary}}\neg\varphi)$ would shed no light on the first query $(\mathrm{B}_{\mathrm{John}}\neg\mathrm{B}_{\mathrm{Mary}}\varphi?)$.

These are extensions we plan to look at in the future.

In the remainder of this paper we will elaborate on this picture. We will explicate the assumptions made about agents, and discuss the multi-dimensional persistence in more detail. The organization is as follows. In section 2 we state the assumptions we make about agents' beliefs, both at single points in time and over periods of time. In section 3 we look closely at persistence in a TBM's with a single datum point. In section 4 we look at TBM's with multiple data points. In section 5 we discuss the extension to data with multiple sequences of agent indices. In section 6 we briefly mention the complexity of the query answering, and in section 7 we briefly mention implementation efforts. We conclude with discussion of related and future work.

2 Assumptions about Belief

We mentioned before that various idealizing assumptions about belief have been made and debated by other researchers, and that the focus of this paper is different from them. Nonetheless a few basic assumptions are essential, and we discuss them here. In the spirit of this paper, we discuss these properties in commonsense terms, rather than in a formal logic.

We have already listed some of the more common restriction on belief: closure of beliefs under tautological implication (as captured by the 'K' axiom), consistency (as captured by the 'D' axiom), and positive and negative introspection (as captured by the '4' and '5' axioms). Since among objective properties (those without a belief operator) we will consider only literals (atomic properties and their negations), the closure property will be irrelevant. Positive and negative introspection will also turn out to impact our results only minimally, as will be discussed in section 5. However, consistency will lie at the heart of the TBM mechanism, and is our first assumption.

Assumption 1 (Consistency) *$\mathrm{B}^t_a\varphi$ and $\mathrm{B}^t_a\neg\varphi$ cannot both hold.*

This is the only assumption we will make about a belief at an instance of time.

In addition we have constraints on how beliefs change over time. We first assume that agents do not come to believe facts without explicitly learning them, but that once they learn them, they do not forget them.

Assumption 2 (Causality and Memory) *If at t agent a does not learn $\neg\varphi$, then $\mathrm{B}^{t+1}_a\varphi$ holds iff $\mathrm{B}^t_a\varphi$ holds.*

Our next assumption is that agents are extremely receptive to new information[Gärdenfors 1988].

Assumption 3 (Gullibility) *If at time t agent a learns φ, then $\mathrm{B}^{t+1}_a\varphi$ holds.*

(Of course, in an environement in which agents are supplied with unreliable or dishonest information, this last assumption would be unacceptable, and we would need a more sophistiated criterion to determine which of the two contradictory facts, the previously believed one and the newly learned one, should dominate.)

Our last assumption is that all these properties are 'common knowledge':

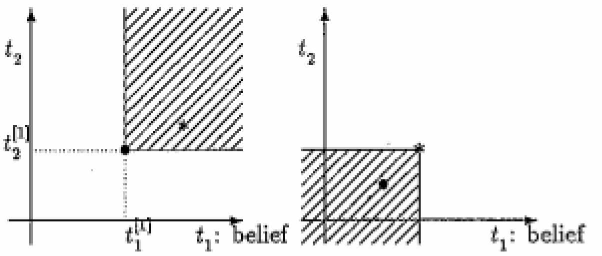

Figure 3: Default region (left) and causal region (right)

Assumption 4 (Common knowledge) *Every agent believes that every agent believes the above properties, that every agent believes that every agent believes them, and so on.*

3 Multi-Dimensional Persistence of a Single Datum

In this section, we consider TBM's induced by a single datum point. We start by considering the non-nested case, in which the datum has the form $\mathrm{L}_a^{t_1^{[1]}} p^{t_2^{[1]}}$ (at time $t_1^{[1]}$ agent a learns that at time $t_2^{[1]}$ property p was (is, will be) true). This induces a 2D TBM, in which the persistences along both axes are uninterrupted and thus do not terminate at all. This situation is represented graphically in Figure 3.

The hatched quarter plane in the left picture, rooted in the point $(t_1^{[1]}, t_2^{[1]})$, is called the *default region* of $(t_1^{[1]}, t_2^{[1]})$. The meaning of this region is that, given only the datum point $\mathrm{L}_a^{t_1^{[1]}} p^{t_2^{[1]}}$, $\mathrm{B}_a^{t_1} p^{t_2}$ holds by default iff (t_1, t_2) lies in that region (i.e., iff $t_1^{[1]} < t_1$ and $t_2^{[1]} < t_2$).

Similarly, if we focus on an affected point (*), all data points affecting it by their forward persistence are distributed in the opposite quarter plane. This is the dual concept of the default region and is called a *causal region* of the affected point. It is depicted graphically in the right picture of the above figure. In this paper we will be concerned mostly default regions.

Finally, although it is only the 2-dimensional case that is so amenable to graphical representation, these concepts extend naturally to the multi-dimensional case. Specifically, given only the datum $\mathrm{L}_{a_1}^{t_1^{[1]}} \cdots \mathrm{L}_{a_{n-1}}^{t_{n-1}^{[1]}} p^{t_n^{[1]}}$, we have that $\mathrm{B}_{a_1}^{t_1} \cdots \mathrm{B}_{a_{n-1}}^{t_{n-1}} p^{t_n}$ holds iff it is the case that $t_1 > t_1^{[1]}, \cdots, t_n > t_n^{[1]}$.

4 Mutiple Data with Incompatible Beliefs

We have so far considered only TBM's induced by a single datum. We now look at the general case in which we have mutiple data. We still assume that all data have the form $\mathrm{L}_{a_1}^{t_1^{[i]}} \cdots \mathrm{L}_{a_{n-1}}^{t_{n-1}^{[i]}} p_i^{t_n^{[i]}}$ or $\mathrm{L}_{a_1}^{t_1^{[j]}} \cdots \mathrm{L}_{a_{n-1}}^{t_{n-1}^{[j]}} \neg p_j^{t_n^{[j]}}$, for some fixed $a_1, \cdots, a_{n-1}$ (again, see section 5 in this connection), but nothing beyond that.

If for any p_k the collection does not contain more than one occurrence of p_k (whether preceded by $\neg$ or not), the situation is simple: the persistence of each fact is independent of the others, and so we construct an independent TBM for each one.

The situation in which multiple occurrence of a p_k exist, but all with the same polarity (that is, either in all data containing p_k the p_k is preceded by $\neg$, or in none), the situation is also simple: the default region is simply the union of the individual regions for each datum containing p_k.

It is the presence of contradictory data that makes the story more interesting. Our assumption of consistency dictates that persistences of contradictory beliefs may not overlap. Without the strong limitations on the form of input data and queries, we would have two problems — to determine which sets of persistences are contradictory, and to resolve the contradiction. For example, we would have to notice that the three sentences $\mathrm{B}_a^t(p \vee q)$, $\mathrm{B}_a^t \neg p$ and $\mathrm{B}_a^t \neg q$ are jointly inconsistent, even though all pairs are consistent. Our restrictions remove this first problem. Since we only consider facts of the form $\mathrm{B}_{a_1}^{t_1} \cdots \mathrm{B}_{a_{n-1}}^{t_{n-1}} p_i^{t_n}$ and $\mathrm{B}_{a_1}^{t_1} \cdots \mathrm{B}_{a_{n-1}}^{t_{n-1}} \neg p_j^{t_n}$, the only fact contradicting $\mathrm{B}_{a_1}^{t_1} \cdots \mathrm{B}_{a_{n-1}}^{t_{n-1}} p_k^{t_n}$ will be $\mathrm{B}_{a_1}^{t_1} \cdots \mathrm{B}_{a_{n-1}}^{t_{n-1}} \neg p_k^{t_n}$, and *vice versa*. When in future work we relax the restrictions on input and queries, we will need a new criterion for determining incompatibility.

Our restrictions do not only render the problem of determining incompatibility trivial, they also simplify the task of resolving it. Since we always have exactly two beliefs contradicting one another, our task reduces to removing one of them; the question is which.[1]

[1] Of course, removing both would also restore consistency, but that would violate our assumption about causality and memory.

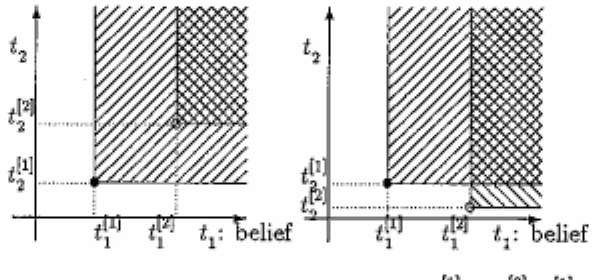

Figure 4: Overlapping default regions ($t_1^{[1]} \neq t_1^{[2]}$, $t_2^{[1]} \neq t_2^{[2]}$)

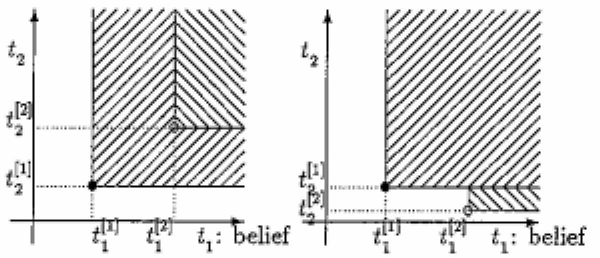

Figure 5: Consistent default regions ($t_1^{[1]} \neq t_1^{[2]}$, $t_2^{[1]} \neq t_2^{[2]}$)

4.1 The 2D Case

The rule for resolving contradictory beliefs in the two dimensional case is derived in a straightforward fashion from the assumptions stated in section 2, and is analogous to the clipping of persistences in simple time maps. We will discuss the case of two data points, but the discussion extends easily to multiple points.

Consider the input data consisting of the two points $\bullet : \mathrm{L}_a^{t_1^{[1]}} p^{t_2^{[1]}}$ and $\circ : \mathrm{L}_a^{t_1^{[2]}} \neg p^{t_2^{[2]}}$. Without loss of generality, assume that $t_1^{[2]} \geq t_1^{[1]}$ holds. We consider the two cases — $t_2^{[2]} \geq t_2^{[1]}$ and $t_2^{[2]} < t_2^{[1]}$ — and assume for now that neither $t_1^{[2]} = t_1^{[1]}$ nor $t_2^{[2]} = t_2^{[1]}$ hold. The default regions of the two points in both cases are shown in the left and right portions of Figure 4, respectively.

In both cases the default regions overlap, which is forbidden, and one of them must be trimmed. In deciding which, we recall the assumption of *gullibility*: right after learning a fact, the agent must believe it. Furthermore, the assumption of memory and causality dictates that the agent must continue to believe it until the next point about which he learns that the fact is false there. This produces the consistent default regions in Figure 5.

Example. If John learns on Monday that on Thursday his house will be painted white ($\bullet$) and on Tuesday he learns that on Friday it will be painted blue (○), then from Monday until Tuesday John will believe that his house will be white from Thursday until the end of time, and from Tuesday on he will believe that his house will be white from Thursday until Friday ($+45^\circ$ shading), and blue afterwards (-45° shading) (the left picture). (Of course, on Thursday he will learn that the painter had a wedding in Chicago and couldn't come.)

On the other hand (the right picture), if John learns on Monday that on Thursday his house will be painted white (•) and on Tuesday he learns that on *Wednesday* it will be painted blue (○), then from Monday until Tuesday John will still believe that his house will be white from Thursday until the end of time ($+45^\circ$ shading), but from Tuesday he will believe that his house will be blue from Wednesday until Thursday (-45° shading), and leave unaltered his belief that it will be white afterwards ($+45^\circ$ shading). (That will change when the painter, back from Chicago a week later, paints John's house turquoise, since neither white nor blue really go well with olive tree in the yard.)

Note that in either case, the beliefs from Tuesday onwards would not change even if the the two pieces of information were acquired in the opposite order. This is no accident; this Church-Rosser property is true in general of our system.

We now turn to the limiting cases, in which either $t_1^{[2]} = t_1^{[1]}$ holds or $t_2^{[2]} = t_2^{[1]}$ holds. Note that from our assumption about the consistency of the input, at most one of them can hold. Therefore, if $t_1^{[2]} = t_1^{[1]}$ holds, we may assume without loss of generality that $t_2^{[2]} > t_2^{[1]}$. This means that at time $t_1^{[1]}(= t_1^{[2]})$ the agent learned that p first became true (•) and later became false (○). The agent will therefore believe at time $t_1^{[1]}(= t_1^{[2]})$ that p will be true from the first point until the second, and false afterwards. There will be nothing later to change that belief, and thus the default region of p forms an infinite horizontal strip, and the default region of $\neg p$ occupies the quadrant above it (Figure 6).

The case in which $t_2^{[2]} = t_2^{[1]}$ holds is more interesting, since it provides insight into the higher dimensional case.

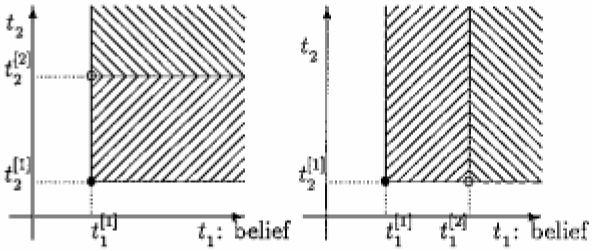

Figure 6: Default regions (left: $t_1^{[1]} = t_1^{[2]}$, right: $t_2^{[1]} = t_2^{[2]}$)

In this case the agent first learned that p became true at some point, and later learned that p became false at that very point. Now in principle we could imagine quite sophisticated criteria to decide which evidence should be given greater credence. However, our assumption of gullibility forces a "recent is better" policy, leading us to accept the later information and abandon the older one. The resulting default regions are shown in the right figure.

4.2 The General Case

We now extend the previous discussion to higher TBM's. We will unfortunately have to do so without the aid of graphics; instead, we will use the following example.

> Example. At $t_1^{[1]}$ you learn that at time $t_2^{[1]}$ your son learned that your son's teacher moved to Japan at time $t_3^{[1]}$ ($L_{you}^{t_1^{[1]}} L_{son}^{t_2^{[1]}} p^{t_3^{[1]}}$). At time $t_1^{[2]}$ you learn that at time $t_2^{[2]}$ your son learned that his teacher moved to the US at time $t_3^{[2]}$ ($L_{you}^{t_1^{[2]}} L_{son}^{t_2^{[2]}} \neg p^{t_3^{[2]}}$) where $t_3^{[2]} > t_3^{[1]}$.
>
> Let $t_1 > \max(t_1^{[1]}, t_1^{[2]})$, $t_2 > \max(t_2^{[1]}, t_2^{[2]})$, and $t_3 > t_3^{[2]}(> t_3^{[1]})$. Then at t_1 you believe that at t_2 your son believes that his teacher is living in the US at t_3. This is true regardless of the relationship between $t_1^{[1]}$ and $t_1^{[2]}$, or the relationship between $t_2^{[1]}$ and $t_2^{[2]}$.
>
> Now consider the same scenario, except that $t_3^{[2]} = t_3^{[1]}$. This means that you believe that your son learned two contradictory facts. However, from the assumption that rules of belief change are common knowledge[2], you know that *your son* will adopt the latest information (as illustrated in the previous figure). Therefore

[2] Note that this is our first use of the common knowledge assumption!

your beliefs about your son's beliefs will depend on the relationship between $t_2^{[1]}$ and $t_2^{[2]}$; if $t_2^{[2]} > t_2^{[1]}$ then you will believe that your son believes that the teacher lives in the US; otherwise you will believe that your son believes that the teacher lives in Japan.

Finally, what will you believe if $t_3^{[2]} = t_3^{[1]}$ and $t_2^{[2]} = t_2^{[1]}$? In this case, *you* will need to break the tie by comparing $t_1^{[1]}$ and $t_1^{[2]}$. Note that they cannot also be equal, as that would violate the assumption that the input data is consistent.

The lesson from this example is clear. To determine whether a point in the hyper-space lies in a particular default region, you should compare the associated time vectors. This ordering is a *reverse lexicographical ordering*, the innermost time being the most significant and the outermost time the least significant.

5 Multiple Sequences of Agent Indices

We have all along assumed one fixed sequence of agent indices in the data: $a_1, \cdots, a_{n-1}$. However, relaxing this limitation is quite simple. Consider data points with multiple sequences of agents indices. Unless we make further assumptions about belief, data with different index sequences will simply not interact. For example, the truth of $B_a^{t_1^{[1]}} B_b^{t_2^{[1]}} p^{t_3^{[1]}}$ is completely independent from the truth of any statement that is not of the form $B_a^{t_1^{[2]}} B_b^{t_2^{[2]}} x$, where x is an objective sentence (containing no belief operator); in particular, it is consistent with $B_a^{t_1^{[1]}} B_c^{t_2^{[1]}} \neg p^{t_3^{[1]}}$. Thus we may simply construct separate TBM's for these different sentences, each obeying our restriction.

However, if we do make further assumptions about belief, we must take greater care. We consider here four possible further assumptions about belief. The first two are the familiar assumptions of introspective capability:

Assumption 5 (Positive introspection) $B_a^t \varphi$ *holds iff* $B_a^t B_a^t \varphi$ *holds.*

Assumption 6 (Negative introspection) $\neg B_a^t \varphi$ *holds iff* $B_a^t \neg B_a^t \varphi$ *holds.*

The other two have to do with beliefs of the agent at different points in time. The first is that not only do agents have memory (which we have already assumed), but they also have perfect memory of past beliefs:

Assumption 7 (Introspection about past beliefs)

$$B_a^t B_a^{t-\tau} \varphi \equiv B_a^{t-\tau} \varphi \qquad \textit{if} \qquad \tau > 0.$$

The last assumption states that agents do not expect their beliefs to change:

Assumption 8 (Belief about stability of beliefs)

$$B_a^t B_a^{t+\tau} \varphi \equiv B_a^t \varphi \qquad \textit{if} \qquad \tau > 0.$$

(Notice that assumptions 5, 7, and 8 can be unified into $B_a^{t_1} B_a^{t_2} \varphi \equiv B_a^{\min(t_1,t_2)} \varphi$.)

We are not arguing on behalf of these assumptions. We list them merely as examples of plausible assumptions one might want to make. The reason we mention them at all is that they violate the property that nested temporal beliefs with different agent indices are independent of one another. For example, under assumption 8, $B_a^2 B_a^3 p^8$ is contradictory with $B_a^2 \neg p^8$.

Fortunately, these four assumptions allow an easy solution. We simply keep simplifying the sentences by substitution, until no further simplifications are possible. It turns out that no matter what subset of these four we choose, the result of this substitution process is unique (the Church-Rosser property again). More generally, whenever our assumptions allow us to derive a unique *canonical form*, we convert the query and the input data to this canonical form, and then revert to our usual procedure. We have not yet investigated the more complex case in which the canonical form is hard to derive or nonexistent.

6 Complexity

Our definition of default regions was constructive, and allows efficient query answering. We briefly discuss the complexity here. If we assume that comparison of a pair of one-dimensional time points is done in one operation, then comparing two n-dimensional time points requires at most n operations. In ordinary applications, n will be a very small integer. Ordinary people will not think of $n = 5$ cases in their everyday life.

If we have N data points, we can get a sorted list of the data points by the priority based on the *reverse lexicographical ordering*, as explained. This requires only $O(n \cdot N \log_2 N) \simeq O(N \log N)$ operations. Since each agent learns informations gradually, it is useful to use a *heap*, a well known balanced tree data structure which can be easily modified to keep ordering.

If we need to identify only the dominant data point in the causal region, even a naive implementation gives it in $O(nN) \simeq O(N)$ operations.

7 Implementation

Our framework can be easily implemented by logic programmming languages such as Prolog as well as ordinary procedural languages such as C. We implemented various versions of this framework in both languages. Backward reasoning mechanism implemented in Prolog employed simplified versions of Kowalski/Sergot's Event Calculus. Forward reasoning mechanism implemented in C employed sorting of an array. As we described before, our algorithm is very fast in simple cases. We intend to implement more complex cases and evaluate their complexity.

As for 2D cases, we have a program which draw a map from a set of data points whose time stamps are given in hour/minutes or minutes/seconds or year/month/day.

Finally, this work has been carried out as part of the research on Agent Oriented Programming. The current simple interpreter, AGENTO[Shoham 1990], only has a simple version of standard time maps. We have implemented an experimental agent interpreter which incorporates the ideas of this paper, and hope to report on it in the future.

8 Related Work and Conclusions

The only closely related work of which we are aware, other than the work on time maps and event calculus which we have discussed at length, is Sripada's [Sripada 1991], which was independently developed. Both systems can deal with nested temporal beliefs. Sripada represents a nested temporal belief by a Cartesian product of time intervals, and like us assumes that nested temporal beliefs are consistent. However, he does not consider the notion of default persistence, and therefore not with the resolution of competing default persistences. It would seem that the result of our system could serve as input to his, but we would like to understand his work better before making stronger claims about the relationship to his work.

As should be clear, much more needs to be done. We made it clear that in this work we did not undertake a logical treatment of time, belief and nonmonotonicity. We were also explicit about the limitations of our framework. We hope to do both in the future, as well as demonstrate the pratical utility of this work.

Acknowledgments

We would like to thank AOP group members at Stanford University and the referees of this paper who gave us useful comments. The first author would like to thank Koichi Furukawa at ICOT, Shigeki Goto, Hirofumi Katsuno, and other colleagues at NTT, too.

References

[Halpern and Moses 1985] J. Y. Halpern, Y. Moses. A Guide to the Modal Logics of Knowledge and Belief: Preliminary Draft, pp.480–490, Proc. of IJCAI, 1985.

[Kowalski and Sergot 1986] R. Kowalski, M. Sergot. A Logic-Based Calculus of Events, New Generation Computing, Vol.4, pp.67–95, 1986.

[Lin and Shoham 1992] F. Lin, Y. Shoham. Persistence of Knowledge and Ignorance (in preparation), 1992.

[McCarthy and Hayes 1969] J. McCarthy, P. J. Hayes. Some Philosophical Problems from the Standpoint of Artificial Intelligence, in B. Meltzer and D. Michie (Eds.), Machine Intelligence 4, Edinburgh University Press, pp.463–502, 1969.

[McDermott 1982] D. V. McDermott. A Temporal Logic for Reasoning about Processes and Plans, Cognitive Science, Vol.6, pp.101–155, 1982.

[Dean and McDermott 1987] T. L. Dean, D. V. McDermott. Temporal Data Base Management, Artificial Intelligence, Vol. 32, pp.1–55, 1987.

[Chellas 1980] B. F. Chellas. Modal Logic, An Introduction, Cambridge University Press, 1980.

[Gärdenfors 1988] P. Gärdenfors, D. Makinson. Revisions of Knowledge Systems Using Epistemic Entrenchment, Proc. of the Second Conference on Theoretical Aspects of Reasoning about Knowledge, pp.83–95, 1988.

[Shoham 1990] Y. Shoham. Agent-Oriented Programming, Stanford Technical Report CS-1335-90, 1990.

[Shoham 1992] Y. Shoham. Nonmonotonic Temporal Reasoning, in D. Gabbay (Ed.), The Handbook of Logic in Artificial Intelligence and Logic Programming (to appear), 1992.

[Sripada 1991] S. M. Sripada. Temporal Reasoning in Deductive Database, PhD thesis of Univ. of London, 1991.

[Allen et al. 1991] J. F. Allen, H. A. Kautz, R. N. Pelavin, J. D. Tenenberg. *Reasoning about Plans*, Morgan Kaufmann Publishers, 1991.

[Hintikka 1962] J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Cornell University Press, 1962.

[Griffiths 1967] A. P. Griffiths ed. *Knowledge and Belief*, Oxford University Press, 1967.

[Konolige 1986] K. Konolige. *A Deduction Model of Belief*, Morgan Kaufmann Publishers, 1986.