Foundations of a Formal Theory of Time Travel

Leora Morgenstern
Courant Institute of Mathematical Sciences
New York University
New York, NY 10012
leora.morgenstern@cs.nyu.edu [*]

Abstract

Although the phenomenon of time travel is common in popular culture, there has been little work in AI on developing a formal theory of time travel. This paper develops such a theory. The paper introduces a branching-time ontology that maintains the classical restriction of forward movement through a temporal tree structure, but permits the representation of paths in which one can perform inferences about time-travel scenarios. Central to the ontology is the notion of an *agent embodiment* whose beliefs are equivalent to those of an agent who has time-traveled from the future. We show how to formalize an example scenario and demonstrate what it means for such a scenario to be motivated with respect to an agent embodiment.

1. Motivation and Overview

The phenomenon of time travel is common in fiction, film, television, and video games. Examples include *A Connecticut Yankee in King Arthur's Court*, , *Slaughterhouse-Five*, *Harry Potter and the Prisoner of Azkaban Back to the Future*, *Bill and Ted's Excellent Adventure*; *Star Trek*, *Lost*, and Chrono Trigger.

This paper explores the development of a formal theory of time travel, in which one would be able to represent and reason with time travel stories.

There are two motivations for this work. First, the growing field of AI and Entertainment has in recent years included research on tools for analysis and generation of video games (see, e.g., (Nelson & Mateas 2008; Whitehead & Young 2009)).

A formal analysis of what makes a time-travel story work—that is, what makes a time-travel narrative coherent, and what inferences in a time-travel scenario are reasonable—could be useful for this line of work.

Second, this is part of a long-term research program in formal models of narrative, and in particular, in developing formal theories that can support determining when narrative structures are coherent. We believe that this is part of the general story-understanding problem, which remains a challenging problem for AI(Louchart, Mehta, & Roberts 2009). Intelligent beings understand time-travel stories (and other science fiction) in much the same way that they understand other works of fiction, which would seem to indicate that it is possible to develop some formal model in which to represent and reason with these stories. In fact, time-travel stories that involve going back to an earlier time in the life of the protagonist may speak to an important theme of literature, used frequently in the works of authors such as Jane Austen (e.g., *Pride and Prejudice*, *Emma*, and *Persuasion*): the reparation of an error made during one's (relative) youth. The desire to correct one's mistakes, to recreate a world in which one can still achieve at least part of one's goals, in which one's missteps are not permanent, is perhaps one of the reasons that time-travel stories have such a hold on our imagination.

Scope: The goal is to develop a representation in which one can represent and reason with time-travel stories. The aim is to develop a formal object-level theory that first, enables representation of a story as a formal *time-travel narrative*, and second, supports the inference that a time-travel narrative is *motivated* with respect to an agent and his goals.

We do not claim that this model explains how time travel might actually happen. That is the focus of physicists' approach to time travel: see Related Work. In this initial work, the bulk of the effort lies in constructing a representation that addresses several conceptual difficulties and showing by example that a time-travel narrative of reasonable complexity can be represented and reasoned with using this structure. This work focuses on backward time travel and restricts time travel to one agent at a time.

Overview and Structure: We first present the working example (adapted from *Star Trek*). Next we discuss an extension to a branching-time temporal ontology that allows representation of paths that correspond to time travel. Central to the ontology is the notion of an *agent embodiment* whose beliefs are equivalent to those of an agent who has time-traveled from the fu-

[*] This paper reports on research done in 2009 and 2010 while the author was visiting research scientist at Courant Institute, and was originally written between June 2009 and January 2010. The author's current address is SAIC, 4001 N. Fairfax Drive, Arlington, VA 22203. She can be reached by email at NYU (leora@cs.nyu.edu) or SAIC (leora.morgenstern@saic.com).

ture. We show how a model can be constructed using this ontology that allows the representation of the working example. Next, we give the formal specification of the ontology and model, give a formal characterization of motivated time-travel narratives, and discuss the inferences that this model supports. We discuss whether time-travel paradoxes arise within the model. We conclude with a discussion of related work, evaluation, and future work.

2.Working Example

The working example in this paper is adapted and simplified (for clarity of presentation) from the Tapestry episode of *Star Trek: The Next Generation.* When Jean-Luc Picard was young, he was involved in a barroom brawl with the Nausicaans, in which he defended the honor of a friend. Picard's heart was irreparably injured in the brawl, and he was given an artificial, damage-prone heart instead. When he is a middle-aged captain of the Enterprise, he is injured. The artifical heart malfunctions, and Picard dies. While he is waiting to enter the afterlife, the superbeing Q explains to him that his premature death is the indirect result of his youthful brawl, and offers him the chance to change his life. Picard returns to the past and avoids the brawl. However, Picard has now become a risk-avoider who never amounts to anything. When Picard realizes the consequences of avloiding the brawl, he asks Q to revert to his old life: he would rather live a meaningful, even if shortened, life.

We develop a model and theory in which we can

• represent Picard's life as it is originally presented;
• represent Picard's life as it would unfold if he avoided the brawl;
• show that it is makes sense for Picard to avoid the brawl after he learns that this has caused his premature death
• show that it makes sense for Picard to change his mind and decide to engage in the brawl once he realizes how his life will unfold if he avoids the brawl.

3. Temporal Ontology

3.1 Representing Backward Time

AI temporal ontologies are typically of two flavors: linear time, as in the event calculus (Miller & Shanahan 1994) and branching time, as in the situation calculus (Reiter 2001). Because linear time does not facilitate reasoning about alternate possibilities, it seems inherently unsuitable as the underlying ontology for time travel. Branching time, with its built-in structure of different possible futures and different possible paths through the temporal tree, seems to have better potential. Still, we are faced with a basic conceptual problem: time goes forward, and not back. The idea is that you can get to any point that exists in the subtree rooted at the point you are now. But in time-travel you want to go back. This is not allowed within standard branching time structures. If you were to draw a link between a point in the branching time structure and a point that temporally precedes it, you would have violated the tree structure. Therefore, a different way must be found.

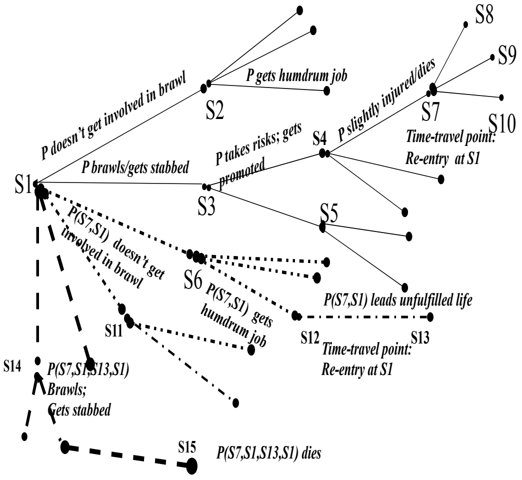
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
S13
S14
S15
P doesn't get involved in brawl
P gets humdrum job
P brawls/gets stabbed
P takes risks; gets promoted
P slightly injured/dies
Time-travel point: Re-entry at S1
P(S7,S1) doesn't get involved in brawl
P(S7,S1) gets humdrum job
P(S7,S1) leads unfulfilled life
Time-travel point: Re-entry at S1
P(S7,S1,S13,S1) Brawls; Gets stabbed
P(S7,S1,S13,S1) dies

Figure 1: *How time travel is effected in a classic forward-branching time tree. Certain time points are designated as time-travel/re-entry points for particular agents. Here S7 is designated as a time-travel point, and S1 is designated as a re-entry point for Picard. This means that an embodiment of Picard moves through a portion of the subtree rooted at S1; the embodiment has full awareness of what happened between S1 and S7. Note also that agent embodiments can be nested. Here, there is a time-travel point along the path taken by the agent embodiment Picard(S7,S1). This results in the nested embodiment Picard(S7,S1,S13,S11)*

We use the following strategies: First, we move away from the implicit assumption in time-travel accounts that there is a path in the time tree that "really" or "first" occurs; and that this path is subsumed when time travel leads to a new path that "really" occurs. In our paradigm, there is no sense of a "real" path through the time tree. Rather, at any point in the time tree, certain sets of paths—more precisely, certain subtrees of the time tree—are accessible by certain embodiments of agents. Some of these subtrees intuitively correspond to how the world might be if time travel were allowed.

Second, we introduce the idea of different *embodiments* of agents. Intuitively, an agent changes as he goes through different experiences, most notably by gaining beliefs. In some sense, one can say that an agent A is different at a later time than at an earlier time. We talk about different embodiments of an agent. Specifically, we talk about the embodiment of an agent $A(Sj, Si)$, where there is a path segment from Si to Sj; this represents an agent who has the memories of an agent who has lived from Si to Sj. Thus, he is aware of what has happened on the path segment from Si to Sj. Intuitively, $A(Sj, Si)$ corresponds to the agent A who has time-traveled back from Sj to Si.

Third, we introduce sets of time-travel (or *choice*) points and re-entry points for particular agents. In general, the path between a re-entry point and a time-travel point is called a *reversible path segment.* There is no specific *action* that takes an agent from a time-travel point to a re-entry point. Rather, an embodiment of an agent will be characterized by the pair (time-travel point, re-entry point), indicating that the agent in this embodiment will have beliefs about what has happened in the reversible path segment. In subtrees that are rooted at re-entry points, there will be paths in which alternate embodiments of agents have active roles. These are the paths which intuitively correspond to time travel.

3.2 Preformal development of the model

Consider the working example, depicted in Figure 1.

At S1, Picard can choose to get involved in the brawl (path from S1 to S3), or to withdraw from the brawl (path from S1 to S2). Picard gets involved in the brawl and is stabbed. Picard continues along the path whose initial segment is (S1,S3). Assume that he is injured again between S4 and S7 and is dead at S7. S7 is a time-travel point: Picard has the choice to "time-travel." The re-entry point available to him is S1, the point at which he decided to get involved in the brawl.

What does time travel in this model actually constitute? As argued above, it would violate the principle of forward branching time to posit an *action* that Picard performs that takes him to S1. Rather, we represent the time travel as a triple consisting of the time-travel point and the re-entry point, together with the agent who is intuitively doing the time-travel along portions of the subtree rooted at S1,

These portions of the subtree are accessible, not to the original Picard who chose to get involved in the brawl, but to a more experienced *embodiment* of Picard. We denote this embodiment as Picard(S7,S1), the Picard who understands what it is like to follow the path from S1 to S7, and chooses now to follow a different path.

However, Picard(S7,S1) can no longer follow the path from S1 to S2. This path was available only to Picard. Picard(S7,S1) can follow a very similar path, consisting of the same action, that of avoiding the barroom brawl. This is depicted in Fig. 1 as the path between S1 and S6. In the formal development, below, we set up an isomorphism between these intuitively similar sets of situations.

Although we never identify any path through the tree as an actual path that happens, in general an embodied agent $A(Sj, Si)$ will at Si have the beliefs of someone who has traversed the path from Si to Sj, much like Scrooge after the visit from the Ghost of Christmas Yet To Come (Dickens 1843).

3.3 Nested Time Travel and Embodiments: This simulation of the time-travel process might be repeated along the path segment of an embodied agent, leading to nested embodiments. The working example provides a simple (and structurally degenerate) example. The embodiment of Picard who is aware that brawling will shorten his life realizes, when middle-aged, that he never amounted to anything because he *avoided* the brawl, and never developed leadership qualities. This embodiment then has a chance to time travel. That is, there is a choice point on this path, subsequent to his realization that he will never amount to anything, and an associated re-entry point at S1.

This is shown in Figure 1. Picard(S7,S1) takes the choice point S13. Therefore, the third embodiment of Picard is denoted Picard(S7,S1)(S13,S1), or more compactly, Picard(S7,S1,S13,S1). In general, the agent embodiment $A(Sj, Si, Sk, Si)$, in which Si occurs more than once, is evidence of a time-travel narrative in which an agent "time travels" to fix a bad choice and later wishes to travel back to the same time point, either to revert to the original choice or to make a still different choice.

Note that this sort of nested embodiment is only one of three ways in which nested embodiments can occur. Two other sorts of nested embodiments can happen when $A(Sj, Si)$ makes an error on the "new" path, and there is a re-entry point prior to the commitment of the error; or when $A(Sj, Si)$ realizes that only a re-entry point prior to Si will be sufficient to fix his affairs.

4. Formal Model

4.1 The Time-Travel Tree Structure

Definition 4.1.1 A time-travel tree structure TTT is a tuple (S,Act,CP,RP,AG, τ, T), whose elements are described below: [1]

S: S is an infinite set of situations, arranged into a partial order under the precedes relation <. The function *time* maps a situation onto its date-clock-time. If $s1 < s2$, then $time(s1) < time(s2)$. There is a path between any two ordered situations. *start* and *end* are functions giving the start and end situations of any finite path segment.

Act: A set of actions of the form $do(ag, ac)$ where ag is an agent (see below) and ac is an *actional*, intuitively, an unanchored action type. *Occurs(do(ag,act),s1,s2)* means that the action of ag performing ac occurs between situations $s1$ and $s2$.

CP: A set $CP \subset S$ of *choice points*, intuitively corresponding to those situations in which an agent can decide to travel to the past. (Or to the future, in models permitting such time travel.)

RP: A set $RP \subset S$ of *re-entry points*, intuitively corresponding to those situations to which an agent time travels.

[1] What follows is based on the theory of knowledge and action in (Davis & Morgenstern 2005); however, belief is used here instead of knowledge.

AG: A set of agent embodiments ag. An agent embodiment (AE) ag may be primary, intuitively an agent who has not (yet) time traveled, or secondary, intuitively one who has time traveled.

A primary AE is represented as a, possibly subscripted; a secondary AE is represented as a $2n+1$-tuple $(a, cp_1, rp_1, \ldots, cp_n, rp_n)$ where a is a primary agent, each $cp_i \in CP$, each $rp_i \in RP$, and (for backward time travel), for each cp_i, rp_i , it is the case that $cp_i > rp_i$.

For $n \geq 1$, we can represent the AE as $(a, cp_1, rp_1, \ldots, cp_{n-1}, rp_{n-1})(cp_n, rp_n)$.The AE $(a, cp_1, rp_1, \ldots, cp_{n-1}, rp_{n-1})$ is the *generating* AE, while $(a, cp_1, rp_1, \ldots, cp_n, rp_n)$ is the *generated* AE. Primary agents can only be generating AEs; secondary agents can be both generating and generated AEs.

Notation 4.1.2: *â* or *a'* is used to range over the primary AE a as well as secondary AEs who are recursively generated by a. This notation is useful when we wish to speak about various embodiments of a specific primary agent. (See the Proof in Example 4.5.2.)

τ: $\tau \subseteq AG \times CP \times RP$. That is, τ is the set of all triples of the form (ag, cp_i, rp_i) which give all the possible ways agent embodiments can travel through the time-travel tree structure. If (ag, cp_i, rp_i) is an element of τ, we say that rp_i is the re-entry point associated with cp_i, from ag's point of view. Note that there may be several re-entry points for a particular choice point of an AE, and several choice points for a re-entry point of a particular AE. Fig. 1 gives an example of the latter scenario.

T: A set of subtrees of S, one for each AE. Assume e is an AE $(a, cp_1, rp_1, \ldots, cp_n, rp_n)$.$T_e$ denotes the subtree rooted at rp_n, the time at which e is first active.

Isomorphisms between subtrees: For each primary AE a, T_a is the subtree of S during which a is active. Let Pa_s denote the subtree of T_a that is rooted at s. If $e = (a, cp_1, rp_1, \ldots, cp_n, rp_n)$, then there is an isomorphism between Pa_s and T_e. Let $\sigma(s)$ denote the image in T_e under this isomorphism. Note that if Pa_s and T_e share the root rp_i, then $\sigma(rp_i) = rp_i$. The existence of this isomorphism is what makes it possible to represent the secondary AE being faced with the same choices that the primary AE faced, and (possibly) making different choices.

4.2 Belief

We use a standard possible worlds semantics of belief as in (Fagin *et al.* 1995) Thus we have the standard definition of belief in terms of belief-accessible worlds:

Definition 4.2.1:
Holds(s,Bel(ag,p)) $\Leftrightarrow$ $\forall$ *s B(ag,s,s')* $\Rightarrow$ *Holds(s',p)*

We need to relate the beliefs of different agent embodiments. For example, it is crucial that Picard(S7,S1) realize that getting involved in the brawl will result in his receiving an artificial heart. If Picard(S7,S1) does not believe this, why should his choice be different from Picard's original choice?

Specifically, we need to be able to say that the more an AE has time traveled—that is, the greater the level of nesting—the more an agent believes. Equivalently, fewer possible worlds are belief-accessible to him.

Stating this axiom is a bit tricky. Recall that a generating AE and the corresponding generated AE inhabit separate worlds. This is built into the structure of the TTT: there are separate, isomorphic structures for generated AEs. Thus it is not the case that the worlds accessible to a generated AE are a *subset* of the worlds accessible to the generating AE. However, we can use the mapping σ between the situations in the isomorphic structures to give us precisely what we need:

Axiom4.2.2:
$B(a(cp_1, rp_1, \ldots, cp_n, rp_n), \sigma(s), \sigma(s'))$
$\Rightarrow B(a(cp_1, rp_1, \ldots, cp_{n-1}, rp_{n-1}), s, s')$

We add the usual KD45 axioms on belief.

To see how Axiom 4.2.2 works, consider the statements P,"If someone gets involved in a barroom brawl, he will have a shortened life span," and Q "If someone avoids a brawl, he becomes a wimp and will not have a meaningful life." Then Picard does not believes either statement in S1; as far as he believes, whatever choice he makes regarding the brawl, he can have both a long and meaningful life. At S7, Picard believes that in any branch in which someone brawls, he will not have a long life. That is, any world which is belief accessible to Picard and which is a successor situation to an AE brawling will be on a path in which the AE has a shortened life. Now, consider all such worlds W, and consider the image of such worlds under the isomorphism σ which maps T_{Picard} to $T_{Picard(S7,S1)}$, denoted $\sigma(W)$. Then, by Axiom 4.2.2, the worlds that are belief accessible to Picard(S7,S1) are a subset of $\sigma(W)$. That is, Picard(S7,S1) believes at least as much as Picard. Therefore, at S1, and in all subsequent situations for Picard(S7,S1), he believes P.

Similarly, at S13, having lived the meaningless wimpy life of the non-brawler, Picard(S7,S1) believes Q; we can show via application of Axiom 4.2.2 that in S1, and in all subsequent situations for Picard(S7,S1,S13,S1), Picard(S7,S1,S13,S1) believes Q as well.

4.3 Time Travel Narratives

We define a time travel narrative (TTN) from the perspective of a primary AE A. Intuitively, a TTN describes the intervals of time through which the AE lives. In our approach, this corresponds to a sequence of path segments in the TTT, with one path segment ending in a choice point, and the next path segment in the sequence beginning with its associated re-entry point.

Defining the TTN is a bit tricky, since a different AE is associated with each path segment. The following notation is helpful: If PS_i is a path segment, then $A(PS_i)$ is the active agent of PS_i.

Definition 4.3.1: A time travel narrative TTN_A is a sequence of path segments $PS_1 \ldots PS_n$ of the TTT that satisfy the following

1. PS_1 is a path segment of T_A.

2. The start and endpoints of the PS_i are characterized recursively as follows:
   (a) The end situation of PS_i is a choice point of A; the starting situation of PS_{i+1} is its associated reentry point.
   (b) For any PS_i, $i > 2$, if $A(PS_{i-1}) = A(cp_1, rp_1, \ldots, cp_{i-2}, rp_{i-2})$, and cp_{i-1} is the end point of PS_{i-1}, then
      i. the starting point of PS_{i-1} is rp_{i-1}, where rp_{i-1} is cp_{i-1}'s associated re-entry point;
      ii. $A(PS_i) = A(cp_1, rp_1, \ldots cp_{i-1}, rp_{i-1})$.

Example 4.3.2: In Fig. 1, TTN_{Picard} is the sequence of path segments ((S1,S7),(S1,S13), (S1,S15)). (S1,S7) represents Picard's involvement in the brawl, leading to his premature death. S7 is a choice point; the associated re-entry point is S1. (S1,S13) represents the path segment in which Picard(S7,S1) avoids the brawl. S13 is the choice point whose associated re-entry point is S1.

4.4 Goals

For any TTN_A, A may have a goal or set of goals. A goal is represented as a fluent. Let G_j be a goal. G_j is achievable iff it holds in some future situation:

Definition 4.4.1:
Holds(s1,Achievable(G_j)*)* $\Leftrightarrow \exists s2 > s1(Holds(s2, G_j))$

We are interested in the cases where it is consistent with an agent's beliefs that a goal is achievable. It is straightforward to show that *Holds(s1,¬ Bel(A, ¬ Achievable*(G_j)*)))* $\iff$
$\exists$ *s2 B(A,S1,s2)* $\wedge \exists s3 > s2(Holds(s3, G_j))$.

A set of goals $G = \{G1 \ldots Gn\}$ is said to be achievable if the conjunction of the goals is achievable. An AE has a set of goals only if it is consistent with his beliefs that the conjunction is achievable:

Axiom 4.4.2: *Holds(s, Goalset(A,G))* $\Rightarrow$
Holds(s,¬Bel(A,¬Achievable$(\bigwedge_{G_j \in G} G_j)$*)))*

Frequently, a goal set is not achievable, leading to the question of which individual goals should be abandoned. We posit an ordering $<_g$ on subsets of G. A preferred subset of G is one that is minimal under this ordering.

4.5 Motivated Time-Travel Narratives

Intuitively, a time-travel narrative is motivated if an AE time travels only when he is in a serious bind and needs to revise history in order to achieve his goals. In this model, this can be expressed by saying that a TTN_A is motivated with respect to the time travel tree if each choice point is taken only after $\hat{A}$ comes to believe that one of his goals can no longer be realized. The associated re-entry point must be chosen so that it is consistent with $\hat{A}$'s beliefs that this life goal, or at least some preferred subset of his goals, can be realized in that re-entry point's future.

Definition 4.5.1
Let TTT be a time-travel tree. Let A be be a primary AE and let $TTN_A = PS_1 \ldots PS_n$. Assume that *Holds(start(PS1), Goalset(G))*. Then TTN_A is motivated with respect to TTT if the following condition holds:

For all $1 < i \leq n-1$,
if *Holds(end(*PS_i*), ¬Bel(A(*PS_i*),Achievable(G)))*, then one of the following is true:
(a) *Holds(start(*PS_{i+1}*),¬Bel(A(*PS_{i+1}*),¬Achievable(G)))*
(b) There is some G' that is a preferred subset of G such that *Holds(end(*PS_i*), ¬Bel(A(*PS_i*),¬Achievable(G')))*
(c) There is some G' that is a preferred subset of G such that
Holds(start(PS_{i+1}*),¬Bel(A(*PS_{i+1}*),¬Achievable(G')))*

Condition (a) holds when it is consistent with one's beliefs that one can achieve all one's goals by starting over (i.e., re-entering the TTT); condition (b) holds when it is consistent with one's beliefs that one's preferred subset of goals can be achieved in the future (thus negating the need to do time travel at all); condition (c) holds when it is consistent with one's beliefs that a preferred subset of goals is achievable at some re-entry point.

Example: One can show that TTN_{Picard} (Example 4.3.2) is motivated with respect to the TTT of the example. We show below a fragment of the axiomatization for this story. (All variables are assumed to be universally quantified with maximum scope unless otherwise specified.)

Fragment of Axiomatization:
Fragment of the causal rules:
(C1) *Occur(do(ag,AvoidBrawl),s1,s2)* $\Rightarrow$
Holds(s,¬ Achievable(MeaningfulLife(ag)))
(C2) *Occurs(do(ag,Brawl),s1,s2)* $\Rightarrow$
Holds(s2, ¬ Achievable(LongLifeSpan(ag)))

Fragment of axioms corresponding to the Tapestry Time Travel Tree:
(N1) *Occurs(do(Picard,Brawl),S1,S3)*
(N2) *Occurs(do(Picard(S7,S1,S13,S1),Brawl),S1,S14)*
(N3) *Holds(S7, ¬ Alive(Picard))*
and so on ...

Fragment of the belief axioms:
(B1) *Holds(S1, ¬Bel(Picard,*
¬Achievable(MeaningfulLife(Picard) $\wedge$
LongLifeSpan(Picard))))
(B2) *Holds(S7, Bel(Picard, C2))*
(B3) *Holds(S13, Bel(Picard(S7,S1), C1))*
(B4) *Holds(s*, Bel(Picard', Occurs(do(Picard',Brawl),S1,s*)*
$\vee$ *Occurs(do(Picard',AvoidBrawl),S1,s*)))*
and so on.

(G1a) ((MeaningfulLife(Picard'), LongLifeSpan(Picard'))
is preferable to ((MeaningfulLife(Picard'))
((MeaningfulLife(Picard')) is preferable to
((LongLifeSpan(Picard'))

The theory includes frame axioms to entail that other fluents do not change value unexpectedly.

Theorem: TTN_{Picard}, is motivated with respect to the time travel tree of Example 4.3.2.

Proof Sketch: At S1, it is consistent with Picard's beliefs that both his goals, having a long life and having a meaningful life, are achievable (B1). At S7, it is clear that Picard's goal of having a long life is no longer achievable (N3). Moreover, Picard now realizes that it is the brawl that caused his inability to achieve a long life: it will always be the case that if he brawls, he will not be able to have a long life (B2,C2). It is still consistent with Picard's beliefs that both goals are achievable (condition a of Def. 4.5.1) as long as his agent embodiment starts over. Picard(S7,S1), like Picard at S7, believes C2 (Axiom 4.2.2); thus Picard(S7,S1) believes that brawling will not allow him to achieve his goals. When Picard(S7,S1) reaches S13, however, he realizes that his goals are not achievable, since he believes both causal rules (C1,B3) as well as (B4). His preference relation causes him to drop the goal of having a long life. (This time, condition c of Def. 4.5.1 is fulfilled.) He returns to S1 with the goal of having a meaningful life. In order to achieve his one life goal, Picard(S7,S1,S13,S1) gets involved in the brawl.

We have shown that at every path segment in which an embodiment of Picard has been active and in which that embodiment comes to realize that his goals are not achievable, one of the conditions of Definition 4.5.1 is satisfied. This proves the theorem.

5. Time-Travel Paradoxes

An advantage of the model proposed here is that the classic paradoxes of time travel do not occur, or occur in a less severe form, within our model.

Time travel is subject to a number of well-known paradoxes and puzzles. Some of the best known are the grandfather/autoinfanticide paradox (Barjavel 1943; Horwich 1987), the predestination paradox (Novikov 1998), and related closed loop and ontological paradoxes.

Autoinfanticide/Grandfather Paradox: In the auto-infanticide paradox (Horwich 1987), an agent travels back to the time when he was an infant, and kills the infant version of himself. (In the similar grandfather paradox, the agent travels back to a time before his parent was conceived, and kills his grandfather.) The difficulty is that at certain points in time — for example, right after the autoinfanticide has occurred — the agent is both alive and dead.

In models of time travel in which one actually follows a spacetime curve to visit the exact world one had lived in previously — that is, in time-travel models espoused by physicists — this can be a serious paradox. One cannot have a world in which one is simultaneously alive and not alive, in which *Holds(s, Alive(A))* and *Holds(s, ¬Alive(A))*.

In the model presented in this paper, the situation is somewhat different. First note that as we have set up the model, generating and generated embodiments cannot act upon one another. (They share situations, but not path segments, in the *TTT*.) Thus, the autoinfanticide scenario could not occur. The grandfather scenario could occur, although we could prevent this by restricting an agent from traveling back to a time before he was born.

Such an approach is unappealing, however, because it requires severe restrictions. Disallowing travel to a time before one was born would disallow such classic time-travel stories as *Connecticut Yankee.* Similarly, we may wish to allow interaction among agent embodiments in order to allow the representation of more time-travel stories. (See the concluding section.)

However, even without these restrictions, there is no real paradox in our model (unlike the physicists' curved space-time model). [2] Specifically, assuming choice point $S2$ and re-entry point $S1$, we do not get any successor situation $S3$ to $S1$ in which *Holds(S3, Alive(A))* and *Holds(S3, ¬(Alive(A)))*. Rather we would get *Holds(S3,Alive(A(S2,S1)))* and *Holds(S3, ¬ Alive(A)))*. This is not a contradiction because generating and generated embodiments are not identical.

Predestination Paradox(Novikov 1998): This occurs when an agent, despite traveling through time, cannot seem to escape a predetermined fate. The agent travels back in time to prevent some action, but by his very actions, causes what he tried to avoid. (The fact that in many stories, this state of affairs is what leads to his backward time travel leads some to label the predestination paradox as a closed loop paradox.) Examples include Boll's *Der Zug War Pünktlicht* and Nesbit's *the Story of the Amulet.* [3] In *The Story of the Amulet*, Jane tries to prevent Caesar from invading England by appealing to the fineness of its culture and citizens; this very description prompts Caesar, who had just decided not to invade, to change his mind and proceed with the invasion.

In physicists' curved spacetime models of closed loop variants of the predestination paradox, such scenarios are troublesome. However, they are not truly paradoxical; they are rather, ungrounded. [4]

In our model, even this notion of ungroundedness is not present. Rather, predestination scenarios occur when sets of branches in the *TTT*, involving all embodiments of a particular primary agent, are constrained so that some set of propositions holds true at some point. A similar analysis holds for the Ontological Paradox.

In fact, it has been argued by philosophers and astrophysicists, such as Horwich and Novikov (Novikov 1998), that the autoinfanticide and grandfather paradoxes are not really paradoxes, since one can ensure that such branches do not occur. It appears to be feasible to formalize this intuition in the model that we are developing. We can specify sets of paths, such as ones

[2] I am indebted to Barbara Partee and Stephanie Lewis for pointing this out.

[3] An example appears as far back as *Oedipus Rex*, though in that case without time travel: An agent, trying to avoid a prophesied and undesirable event, performs actions which in fact lead to the event happening.

[4] Compare, e.g., the account of (Kripke 1975) contrasting truthteller and liar sentences. While truthteller sentences are ungrounded,they are not paradoxical.

in which an embodiment of an agent kills the agent's grandfather before the agent's parent is born, as incoherent. Or we can be sure to set up the action structure so that certain actions are not feasible for certain agent embodiments.

While the predestination paradox appears puzzling and is certainly frustrating to the protagonist of a story, it is not really a paradox at all. Indeed, it may be desirable to set up one's time tree so that certain sets of paths are constrained to entail a particular fact (e.g., the occurrence of a particular event). Our model supports and encourage this phenomenon.

5. Related Work

Modern theories of narratology, such as (Bal 2009) and (Abbot 2008), discuss representations of non-standard time within narratives. However, their focus has generally been on narrative constructs such as flashbacks, or the deliberate abuse of text order in representing time, as in Martin Amis's *Time's Arrow.* We know of no discussion of time travel. In addition, none of the work in this field attempts to develop formal theories in which a story can be represented and reasoned about.

The interest of physicists in time travel dates back to the development of Einstein's theory of relativity, and his contention that spacetime is locally curved. Examples of discussions of time travel's feasibility include (Gödel 1949; Malament 1985; Friedman *et al.* 1990).

It should be noted that the focus of these works is different from the work described in this paper. Physicists are primarily interested in demonstrating that the space-time continuum can curve back upon itself, allowing a person in certain circumstances to travel back to a time that he had lived through before. The focus of their work is in formalizing the topological features of spacetime structures, such as closed spacetime curves and wormholes, which enable time travel. Topology aids as well in the analysis of time-travel paradoxes.

In contrast, we have developed a model in which there are no spacetime curves that go back on themselves. Rather, we extend a classical forward-branching time tree using the notion of agent embodiments, and use these to represent paths that simulate time travel.

Some of the ideas that are foundational to this work, such as time travel spawning the existence of many versions of a person, with varying amounts of knowledge, are prevalent in popular culture. See., e.g., the discussion in (Wagland 2007). The contribution in this paper has been to formalize this notion, and specifically, to construct a temporal structure that enables a consistent formalization.

6. Preliminary Evaluation, Future Work

The goal of this research project is the development of a formal theory of time travel that enables representing and reasoning with time-travel stories. We have shown that the model introduced in this paper can represent the Star Trek Tapestry episode. The larger questions are: How general is this theory? What other time travel stories can it handle? What is outside of its scope?

As a first step in the evaluation of this theory, we classified several dozen time-travel stories (from fiction, film, and TV) with regard to several salient features of time travel. For example, some time-travel accounts allow time travel only within an agent's lifetime; some accounts restrict time-travel to a single agent.

Next, we did a preliminary analysis of the ability of the theory to represent this sort of time-travel story, assigning to each class one of 5 rankings, depending on whether the theory can or cannot handle the time-travel story, or requires minor, moderate, or major changes to handle the story. A portion of the table is shown below.

Feature	Example in Fiction	Handles?
Travel to past, within lifetime	Christmas Carol	yes
Travel to any time in past	*Conn. Yankee*	yes
Future time travel	*Time Machine*	minor
Agent travels with object	*Story of the Amulet*	yes
Agent embodiment inhabits self	Star Trek Tapestry	yes
Agent embod. observes self	*Prisoner of Azkaban*	moderate
Agent emobd. interacts w. self	*Back to the Future2*	moderate
Multi-agent time travel	*Wizards of Waverly Place*	major
One predetermined future	*Slaughterhouse-Five*	no

Our analysis suggests that the theory as thus far developed can handle a reasonably large class of time-travel stories. At the same time, the analysis also suggests next steps in this research, namely, making modifications needed in order to enable representation of the time-travel stories that we cannot currently handle. A first step is extending this work to future time travel. (Note, however, that in our model, there is no concept of "the future"; one travels to *a* future, not *the* future.) We then plan to modify the model and theory to allow interaction between generating and generated agent embodiments.

Future work also includes using the insights developed during this research to develop a formal theory of narratives of mistake reparation. We plan to explore the connection between re-entry points in time-travel narratives and fictional situations that are propitious for a character to re-apply himself to his original goals. We are interested both in situations where these propitious situations are serendipitous (as in*Pride and Prejudice* and *Persuasion*) and those in which the propitious situations are explicitly planned for and engineered by the protagonist (as in *Emma*).

Acknowledgements: This paper was written in 2009 and 2010 while I was visiting research scientist at Courant Institute of Mathematical Sciences. Many of the ideas were developed in talks I gave at a seminar at the University of Toronto in June 2009; the LSIR2 Workshop at IJCAI in July 2009; and Richmond Thomason's festschrift at the University of Michigan in November 2009. I am grateful to Ernie Davis, Hojjat Ghaderi, Gerhard Lakemeyer, Stephanie Lewis, Sheila McIlraith, and Barbara Partee for their helpful comments and suggestions.

References

Abbot, H. P. 2008. *The Cambridge Introduction to Narrative, 2nd Edition.* Cambridge: Cambridge University Press.

Amis, M. 1991. *Time's Arrow.* England: Jonathan Cape.

Bal, M. 2009. *Narratology: Introduction to the Theory of Narrative, 3rd Edition.* Toronto: University of Toronto Press.

Barjavel, R. 1943. *Le Voyageuer Imprudent.* Gallimard.

Böll, H. 1949. *Der Zug war pünktlich.* Germany: Friedrich Middelhauve.

Davis, E., and Morgenstern, L. 2005. A first-order theory of communication and multi-agent plans. *Journal of Logic and Computation* 15(5):701–749.

Dickens, C. 1843. *A Christmas Carol.* England: Chapman and Hall.

Fagin, R.; Halpern, J. Y.; Moses, Y.; and Vardi, M. Y. 1995. *Reasoning About Knowledge.* Cambridge, Massachusetts: The MIT Press.

Friedman, J.; Morris, M. S.; Novikov, I. D.; Fernando, E.; Klinkhammer, G.; Thorne, K. S.; and Yurtsever, U. 1990. Cauchy problem in spacetimes with closed timelike curves. *Physical Review D* 42(6):1915–1930.

Gödel, K. 1949. An example of a new type of cosmological solution of Einstein's field equations of gravitation. *Reviews of Modern Physics* 21:447–450.

Heinlein, R. A. 1941. By his bootstraps. *Astounding Science Fiction.* Published under pen name Anson MacDonald.

Horwich, P. 1987. *Asymmetries in Time.* MIT Press.

Kripke, S. 1975. Outline of a theory of truth. *Journal of Philosophy* 19:690–716.

Lewis, D. 1976. Paradoxes of time travel. *American Philosophical Quarterly* 145–152.

Louchart, S.; Mehta, M.; and Roberts, D. L., eds. 2009. *Intelligent Narrative Technologies II*, volume SS-09-06 of *AAAI Spring Symposium Series.*

Malament, D. 1985. Minimal acceleration requirements for "time travel" in Gödel space-time. *Journal of Mathematical Physics* 26(4):774–777.

Miller, R., and Shanahan, M. 1994. Narratives in the situation calculus. *J. Log. Comput.* 4(5):513–530.

Nelson, M. J., and Mateas, M. 2008. An interactive game-design assistant. In *IUI*, 90–98.

Nesbit, E. 1906. *The Story of the Amulet.* England: T. Fisher Unwin.

Novikov, I. 1998. *The River of Time.* Cambridge U. Press.

Reiter, R. 2001. *Knowledge in Action.* Cambridge, Masachusetts: MIT Press.

Rowling, J. 1999. *Harry Potter and the Prisoner of Azkaban.* Bloomsbury.

Twain, M. 1889. *A Connecticut Yankee in King Arthur's Court.* Charles L. Webster and Co.

Vonnegut, K. 1969. *Slaughterhouse-Five.* New York: Dell.

Wagland, S. 2007. Time travel and resolving paradoxes in fiction. http://www.soundedit.com.au/swaggers/paradox.html.

Wells, H. G. 1895. *The Time Machine.* England: William Heinemann.

Whitehead, J., and Young, R. M., eds. 2009. *Proceedings of the 4th International Conference on Foundations of Digital Games, FDG 2009, Orlando, Florida, USA, April 26-30, 2009.* ACM ACM.