# Representing and Reasoning about Time Travel Narratives: Foundational Concepts

**Leora Morgenstern**
Leidos Corporation
Arlington, VA 22203
leora.morgenstern@leidos.com

**Abstract**

The paper develops 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.

## 1 Introduction

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 narratives. This is part of a long-term research program in developing formal models of narrative (Morgenstern 2008). A longer version of this paper is available at (Morgenstern 2014).

## 2 Working Example

Our working example is adapted and simplified from the Tapestry episode of *Star Trek: The Next Generation*. When Jean-Luc Picard was young, he was involved in a brawl with the Nausicaans, in which he defended the honor of a friend. Picard's heart was irreparably injured, 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 offers him the chance to change his life. Picard returns to the past and avoids the brawl. However, Picard has now become overly cautious and thus 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.

## 3 Temporal Ontology

### 3.1 Representing Backward Time

AI temporal ontologies are generally founded on linear time, as in the event calculus (Miller and Shanahan 1994) or branching time, as in the situation calculus (Reiter 2001). Because linear time does not facilitate reasoning about alternate possibilities, it seems unsuitable as the underlying ontology for time travel. Branching time seems to have better potential. but we are left with a basic conceptual problem: time goes forward, not back.

To solve this problem, we first move away from the implicit assumption in time-travel accounts that there is a path in the time tree that "really" or "first" occurs; Rather, at any point in the time tree, certain sets of paths are accessible by certain 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. Specifically, we talk about the embodiment of an agent $A(Sj, Si)$, where there is a path segment from $Si$ to $Sj$; this represents someone who has the memories of an agent who has lived from $Si$ to $Sj$. $A(Sj, Si)$ corresponds to the agent $A$ who has time-traveled back from $Sj$ to $Si$.

Third, we introduce sets of time-travel (or *departure*) points and re-entry points for particular agents. In general, the path between a re-entry point and a departure point is called a *reversible path segment*. An embodiment of an agent will be characterized by the pair (departurel point, re-entry point), indicating that the agent in this embodiment will have beliefs about what has happened in the reversible path segment.

### 3.2 Preformal development of the model

Consider the working example (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.

We represent time travel as a triple consisting of the departure 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 to a more experienced *embodiment* of Picard, Picard(S7,S1), who

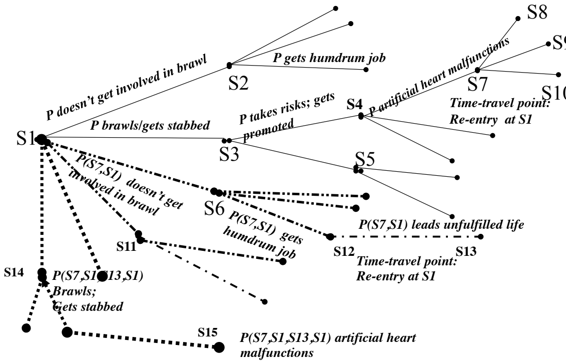


Figure 1: *Star Trek TNG Tapestry Example.*

chooses now to follow a different path. He can no longer follow the path from S1 to S2 but he follows a similar path in which he avoids the brawl (S1,S6).

**Nested Time Travel and Embodiments:** The time-travel process can be repeated along the path segment of an embodied agent, leading to nested embodiments. E.g., 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. This embodiment then has a possibility of time travel at S13 and an associated re-entry point at S1. (See Fig. 1.)

Embodiment along a path stays the same or grows deeper, but does not grow shallower. An agent is marked by the path segments that he believes he has lived.

# 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,DP,RP,AG, $\tau$, 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 $s2$.

**DP:** A set *DP* $\subset$ *S* of *departure points*, intuitively corresponding to those situations in which an agent can decide to travel to the past, or is involuntarily dispatched to a point inthe 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.

**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 $(a, dp_1, rp_1, \ldots, dp_n, rp_n)$ where $a$ is a primary agent, each $dp_i \in DP$, each $rp_i \in RP$, and (for backward time travel), for each $dp_i, rp_i$ , it is the case that $dp_i > rp_i$.

For $n \geq 1$, we can represent the AE as $(a, dp_1, rp_1, \ldots, dp_{n-1}, rp_{n-1})(dp_n, rp_n)$.The AE $(a, dp_1, rp_1, \ldots, dp_{n-1}, rp_{n-1})$ is the *generating* AE, while $(a, dp_1, rp_1, \ldots, dp_n, rp_n)$ is the *generated* AE. Primary agents can only be generating AEs; secondary agents can be both generating and generated AEs.

**Notation 4.1.2:** $\hat{a}$ 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.)

**$\tau$:** $\tau \subseteq AG \times DP \times RP$. That is, $\tau$ is the set of all triples of the form $(ag, dp_i, rp_i)$ which give all the possible ways agent embodiments can travel through the time-travel tree structure. If $(ag, dp_i, rp_i)$ is an element of $\tau$, we say that $rp_i$ is the re-entry point associated with $dp_i$, from $ag$'s point of view. Note that there may be several re-entry points for a particular departure point of an AE, and several departure 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 $(a, dp_1, rp_1, \ldots, dp_n, rp_n)$.$T_e$ 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 $e$ = $(a, dp_1, rp_1, \ldots, dp_n, rp_n)$, 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 be able to say that the more an AE has time traveled, the more an agent believes. Equivalently, fewer possible worlds are belief-accessible to him. 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. We can use the mapping $\sigma$ between the situations in the isomorphic structures to give us precisely what we need:

**Axiom4.2.2:**
$B(a(dp_1, rp_1, \ldots, dp_n, rp_n), \sigma(s), \sigma(s'))$
$\Rightarrow B(a(dp_1, rp_1, \ldots, dp_{n-1}, rp_{n-1}), s, s')$

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

We add the usual KD45 axioms on belief.

E.g., 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, 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 $\sigma$ 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 departure 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 departure point of $A$; the starting situation of $PS_{i+1}$ is its associated re-entry point.
   (b) For any $PS_i$, $i > 2$, if $A(PS_{i-1}) = A(dp_1, rp_1, \ldots, dp_{i-2}, rp_{i-2})$, and $dp_{i-1}$ 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 $dp_{i-1}$'s associated re-entry point;
      ii. $A(PS_i) = A(dp_1, rp_1, \ldots dp_{i-1}, rp_{i-1})$.

**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 departure 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 departure 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 departure 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. Proof in (Morgenstern 2014).

## 5 Supplementary Examples

In the full paper (Morgenstern 2014), we distinguish between voluntary and involuntary time travel, and give formal characterizations of Second Chance narratives, in which a protagonist has the chance to fix a mistake in his past, and Serendipitous Nnarratives, in which a protagonist unexpectedly has a chance to improve his chances or circumstances. Examples of involuntary time travel are in *Peggy Sue Got Married* and *Groundhog Day*; *Peggy Sue* is a Second Chance narrative, while *Groundhog Day* is a Serendipitous narrative. Only some serendipitous narratives are time-travel stories; examples that are not time-travel narratives include many of the novels of Jane Austen, including *Persuasion*, *Emma*, and *Pride and Prejudice*.

## 6 Time-Travel Paradoxes

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), in which an agent travels back to the time when he was an infant, and kills his grandfather or the infant version of himself; the predestination paradox (Novikov 1998), in which no matter what an agent does, he cannot fix the past to prevent some disaster occurring; and closed loop and ontological paradoxes. In the full paper, we show that an advantage of the model proposed here is that these classic paradoxes of time travel do not occur, or occur in a less severe form, within our model.

## 7 Related Work

Modern theories of narratology, such as (Bal 2009) and (Abbot 2008), discuss representations of non-standard time within narratives. Their focus has generally been on narrative constructs such as flashbacks. 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).

## 8 Evaluation and Future Work

We have classified dozens of time-travel stories (from fiction, film, and TV) with regard to several features and have evaluated whether our theory can handle such features and/or the extent of modification that would be necessary. A sample of this analysis is show in figure 2.

Future work includes extending the theory to handle more cases, especially that of interacting time-traveling agents, and exploring the connection between time-travel and Second Chance narratives.

| Feature | Example in Fiction | Handles? |
|---|---|---|
| Travel to past within lifetime | *Christmas Carol, Peggy Sue, Groundhog* | 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 embodiment observes self | *Prisoner of Azkaban* | moderate |
| Agent emobdiment interacts w. self | *Back to the Future2* | moderate |
| Multi-agent time travel | *Wizards of Waverly Place* | major |
| One predetermined future | *Slaughterhouse-Five* | no |
| Agent cannot change future or past but can bring knowledge back and forth | *Time Traveler's Wife* | no |

Figure 2:

### Acknowledgements

I am grateful to Ernie Davis, Hojjat Ghaderi, Gerhard Lakemeyer, Stephanie Lewis, Sheila McIlraith, Barbara Partee, and Robert Sloan for their helpful comments and suggestions,and to MZ for his indefatigable reading and viewing of hundreds of time-travel stories and episodes, which has helped greatly with this research.

## References

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

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

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

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.

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.

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

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.

Morgenstern, L. 2008. A first-order theory of stanislavskian scene analysis. In *AAAI 2008*. AAAI Press.

Morgenstern, L. 2014. Representing and reasoning about time travel narratives: Extended paper. http://www-formal.stanford.edu/leora/papers or in ResearchGate.

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

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