Proceedings, The Thirteenth AAAI Conference on
Artificial Intelligence and Interactive Digital Entertainment (AIIDE-17)

Causal Link Semantics for Narrative Planning Using Numeric Fluents

Rachelyn Farrell, Stephen G. Ware
rfarrell@uno.edu, sgware@uno.edu
Narrative Intelligence Lab, Department of Computer Science
University of New Orleans, New Orleans, LA, 70148, USA

Abstract

Narrative planners would be able to represent richer, more realistic story domains if they could use numeric variables for certain properties of objects, such as money, age, temperature, etc. Modern state-space narrative planners make use of causal links—structures that represent causal dependencies between actions—but there is no established model of a causal link that applies to actions with numeric preconditions and effects. In order to develop a semantic definition for causal links that handles numeric fluents and is consistent with the human understanding of causality, we designed and conducted a user study to highlight how humans perceive enablement when dealing with money. Based on our evaluation, we present a causal semantics for intentional planning with numeric fluents, as well as an algorithm for generating the set of causal links identified by our model from a narrative plan.

Introduction

AI planning has proven a popular formalism for representing and generating narratives (Young et al. 2013). A story world is encoded as a problem domain specifying all its characters, places, objects, and possible actions. Given a description of the initial state and the author's goal, a narrative planner searches the space of possibilities for a series of actions that achieves the author's goal while ensuring that every character has a reason for his or her actions.

State-of-the-art narrative planners either use propositional predicate logic, where every proposition can either be true or false, or multi-valued variables, which can have one of a finite set of possible values (Helmert 2006). These representations fall short of capturing an important element of the real world—the notion of *quantity*. With propositional predicates we can say that the character Bob possesses some gold by setting the predicate *hasGold*(Bob) to True. With multi-valued variables we can say that there is a Gold object whose location is Bob by setting the value of the fluent *location*(Gold) to Bob. In this case, we could represent how much gold Bob has by encoding each of the different pieces of gold as individual objects, and then counting the ones whose location is Bob, but this is often an unnecessary amount of work. There is no additional information gained by keeping track of which pieces of gold Bob has.

It is much more efficient, and more intuitive, to represent the total amount of gold Bob possesses as a single *numeric* fluent, e.g. *gold*(Bob). Effects of actions can modify the fluent's value, and preconditions can check it against any logical condition. Many classical planners have utilized numeric fluents since they were adopted into the Planning Domain Definition Language (PDDL) in 2003 (Fox and Long 2003). However, narrative planners based on those algorithms have been unable to utilize this representation because of a complication that arises when it is combined with *causal links*.

Causal links represent causal dependencies between actions. Narrative planners use them primarily to keep track of how characters intend to achieve their goals. In general, if the effect of an action is used to satisfy a later action's precondition, then a causal link is drawn from the former action to the latter. When using propositional predicates, the question of whether a link should be drawn is straightforward: if the effect of an action is some proposition p, and the precondition of a later action is p, and there are no actions between them which have the effect $\neg p$, then a link should be drawn. When we replace propositional or multi-valued variables with numeric fluents, it becomes less obvious.

Imagine this scenario: Bob currently has no gold, and then he takes an action that earns him 1 gold. Afterwards, he buys something that costs 1 gold. We can most likely agree that a link should be drawn from the *earn* action to the *buy* action, because the former clearly enabled the latter; it would have been impossible for Bob to buy the item had he not first earned the 1 gold. However, what happens when we expand this example to incorporate more actions and more gold?

Imagine if Bob had taken four different actions that each earned him 1 gold, and then bought an item that cost him 2 gold. Now, which of the four *earn* actions should be linked to the *buy* action? Perhaps all of them, since they all contributed to his total amount of gold. Or perhaps just the second one, because that's the one that caused the precondition of the *buy* action, ($gold$(Bob) ≥ 2), to become true. Or perhaps just the first two, because it took *two* gold to enable the precondition. Or maybe just the last two, because those are the two most *recent* actions that can account for the two gold he needed. It becomes even more complicated when we introduce additional *buy* actions... which *earns* should be linked to which *buys*?

To be clear, we are not presenting a causal link planner.

Our motivation is to develop an extension for state-space intentional narrative planners—e.g. Glaive (Ware and Young 2014)—that provides support for numeric fluents. Because these planners utilize causal links to reason about character intentions, we require a model of causality that accurately represents how humans perceive enablement between actions with numeric preconditions and effects. We designed and conducted a user study to highlight how humans understand causality when dealing with money. Based on our evaluation, we present a causal semantics for intentional planning with numeric fluents, as well as an algorithm for generating the set of causal links identified by our model from a narrative plan.

Related Work

Causal Links

Causal links were introduced for Partial-Order Causal Link (POCL) planners, a family of planners that searches a space of possible plans, rather than of possible world states (McAllester and Rosenblitt 1991). Each node in plan-space represents a partial, or incomplete plan, and the edges between them represent additions or fixes to the plan. A *causal link* represents a commitment that an effect of an earlier step (the tail) will be used to satisfy a precondition of a later step (the head). POCL planners guarantee this commitment by not allowing any step that undoes the effect to be ordered between the tail and the head of a causal link.

Definition: A causal link $s \xrightarrow{p} t$ exists from event s to event t for proposition p iff s occurs before t, s has an effect p, t has a precondition p, and no event occurs between s and t which has the effect $\neg p$. We say s is the *causal parent* of t, and that an event's *causal ancestors* are those events in the transitive closure of this relationship.

Causal relatedness between narrative events plays a significant role in how humans comprehend stories (Trabasso and Sperry 1985; Trabasso and Van Den Broek 1985). POCL plans were adapted for use in story generation, due in part to their inherent causal structure (Young 1999). This gave researchers the opportunity to explore other narrative uses for causal links.

Notably, they were used to represent *frames of commitment* for different characters in the IPOCL narrative planner, which added the notion of *intentionality* (Riedl and Young 2010). In classical planning, a plan is valid iff every action's preconditions are true at the time it is executed and the goal is true at the end. For an IPOCL plan to be valid, there is an additional constraint: For every action, for every character who takes that action, that action must either achieve one of that character's goals or be the *causal ancestor* of such an action. This improves character believability by ensuring that actions are properly motivated and goal-oriented for the characters who take them.

This model of intentional narrative planning was adopted for the Glaive Narrative Planning System, which uses modern fast state-space heuristic search techniques but retains the knowledge representation afforded by causal links (Ware and Young 2014). In addition to character intentionality, causal links have been used to model narrative discourse techniques including conflict (Ware et al. 2014), emotion (Gratch 2000), and event salience (Cardona-Rivera et al. 2012).

paintPortrait(?char)
Precon: {}
Effects: money(?char) += 300

paintLandscape(?char)
Precon: {}
Effects: money(?char) += 100

payBills(?char)
Precon: money(?char) >= 100
Effects: money(?char) -= 100

buyTV(?char)
Precon: money(?char) >= 300
Effects: money(?char) -= 300

getRobbed(?char)
Precon: money(?char) >= 100
Effects: money(?char) -= 100

Figure 1: Commissions Domain

Numeric Fluents

Numeric fluents were adopted into the Planning Domain Definition Language (PDDL) in version 2.1 in an effort to support more realistic planning domains (Fox and Long 2003). Since then, there have been many planners to successfully handle numeric fluents (Gerevini, Saetti, and Serina 2008; Coles et al. 2012; Eyerich, Mattmuller, and Roger 2012). However, by that time the planning community had mostly moved away from partial-order planners to state-space planners that no longer required causal links. To our knowledge, no one has yet addressed the specific challenge of using numeric fluents in conjunction with causal links.

Models

We would like a model for how to answer this question: Given a plan and an action in that plan with a numeric precondition, which step(s) in the plan *enabled* the action by contributing to that precondition?

In the traditional definition of a causal link, the tail step *satisfies* the precondition p of the head step. We might be tempted to consider a very direct translation of this concept: link only one tail step for a single precondition of the head step; namely, the one that finally *satisfies* the condition, making it become true when it was previously false. Using the Commissions domain displayed in Figure 1, consider the example plan below (Figure 2):

After the first *paintLandscape* action, the character's money total is 100. After the second, it is 200. Only after the third action does it become 300, thus satisfying the precondition of the *buy* action (money ≥ 300). Therefore, if we use the model mentioned above, only the third *paintLandscape* will be causally linked to the final *buyTV*. However, the first

Action	Effect	Money
(initial)	—	0
paintLandscape	+100	100
paintLandscape	+100	200
paintLandscape	+100	300
buyTV	-300	0

Figure 2: Example Problem

two *paintLandscape* actions are critical in this story; if we were to remove either of them from the plan, the plan would become impossible.

A better answer for this example would be to link all three of the *paintLandscape* actions to *buyTV*. A model that could achieve this might be: For an action with a precondition involving the numeric variable *v*, link all steps prior to it whose effects modify *v* in the appropriate direction. (If the precondition requires the variable to be sufficiently *large,* then we consider only the steps whose effects *increase* it, and ignore those whose effects *decrease* it.) This would yield the desired result for the example above; all three *paintLandscape* actions would be linked. However, if there were ten *paintLandscape* actions instead of three, all ten would be linked to *buyTV* even though only three of them were actually necessary to achieve the amount required by the precondition.

Although this conservative approach may in fact be more in line with how some humans perceive causality in this scenario, it causes a problem for intentional planners; it allows actions to be causally linked to goals that were already enabled prior to the action taking place. This means that the planner would *explain* the character's willingness to paint seven more landscapes by saying that she did it to contribute to her goal of having a TV. Yet, in the story, she already had enough money to buy the TV, so her goal was already enabled. This could lead to a loss in character believability—characters taking actions that the audience perceives as unlikely or not properly motivated.

In summary, we seek to identify the best model for determining where causal links should be drawn and where they should not be drawn. We would like our model to adhere to the traditional definition of a causal link in that it represents a causal *dependency*: Removing a step which is the tail of a causal link should invalidate the plan. We would also like to preserve the aforementioned *intentionality* constraint—that actions should not be linked to goals which, at the time of execution, were already enabled. Finally, we would like our model to reflect humans' understanding of enablement as accurately as possible.

We begin by identifying features of a possible model and dividing these features into two categories: those which answer the question of when to *start* counting links, and those which answer the question of when to *stop* counting them. In all cases, we start from the head step (in our example, the *buyTV* action) and work backwards to consider each action that modifies the critical variable in the appropriate direction. There are two options for when a model could start:

Step #	Action	Effect	Money
0	(initial)	—	0
1	paintLandscape	+100	100
2	payBills	-100	0
3	paintLandscape	+100	100
4	payBills	-100	0
5	paintLandscape	+100	100
6	paintLandscape	+100	200
7	paintLandscape	+100	300
8	payBills	-100	200
9	paintLandscape	+100	300
10	paintLandscape	+100	400
11	buyTV	-300	100

Figure 3: Example Problem

IMM — Start immediately, i.e. always add a link for the first candidate action we come to.

SKP — Skip any action for which the target precondition was already true in the state prior to its execution.

Consider the problem in Figure 3. We want to answer the question, "Which action(s) enabled the *buyTV* action?" If we are using the IMM option, then the first causal link will be for Step 10—the most recent *paint* action prior to *buyTV.* If we are using the SKP option, then we will skip this step because the precondition (money $\geq$ 300) was already true before it was executed. When we consider Step 9, we see that the value of *money* was only 200 prior to its execution, so we cannot skip this step. Therefore, the first link will be for Step 9.

To determine when to stop, we have five possibilities:

ONE — Stop after linking one action.

ACC — Stop after linking enough actions to *account for* the amount required by the precondition. In the example in Figure 2, the precondition requires *money* to be at least $300, so this option would cause us to stop after *money* has been increased by $300—either Steps 10, 9, and 7; or Steps 9, 7, and 6 (depending on where we started).

ACC+ — Stop after linking enough actions to account for the precondition AND any actions that change the variable in the opposite direction, if they occur before we have accounted for the precondition. In this domain, the *payBills* actions cause the value of *money* to decrease by $100, so for every *payBills* action we come across, we increase the value we need to account for by 100. For example, if we start with Step 10, then we would link Steps 10, 9, 7, and 6; because Step 8

caused us to need to account for $100 extra. After linking Step 6, we have now accounted for the total $400, so we stop.

ACC++ Stop after linking enough actions to account for the precondition and any actions that change the variable in the opposite direction, if they occur *anywhere* in the story prior to the head step. In this example, there are a total of three *payBills* actions prior to *buyTV*, so we must account for $300 extra—a total of $600 dollars. Thus, if we started with Step 10, we would link Steps 10, 9, 7, 6, 5, and 3.

ALL Stop only upon reaching the initial state, i.e. link *all* actions prior to the head step that modify the variable in the right direction. This time, if we started with Step 10, we would link Steps 10, 9, 7, 6, 5, 3, and 1.

We considered all possible combinations of when to start and when to stop, for a total of 10 possible models.

Evaluation

The purpose of this experiment was to solicit from humans, using natural language, where the causal links should be drawn (and by extension, where they should not be drawn). We chose to use money as our numeric variable because people are familiar with it and are comfortable reasoning about earning and spending.

We designed a set of seven stories based on the Commissions domain (Figure 1) to target each of the features outlined in the previous section. Participants were recruited using Amazon Mechanical Turk and were shown all seven stories. In each story the character Jessica starts out with $0, and then a series of actions occurs. For every action in which she earns money (the *paintLandscape* and *paintPortrait* actions), participants were asked why Jessica took that action. They chose an answer from the following list:

- So she could buy a TV
- So she could pay her bills
- So she could buy a TV *and* so she could pay her bills
- So she could buy a car
- None of the above

Participants' answers reveal where they believe the *outgoing* links from the action in question should go. That is, each *paint* action can be linked either to *buyTV*, to *payBills*, to both *buyTV* and *payBills*, or have no outgoing links. The *buy a car* option was included as a quality control filter (discussed below). Once we established the "correct" answers as identified by humans, we used each of our ten models to generate links for the same set of stories, and compared them to determine which models were more successful at capturing the human interpretation of causality. It is important not to confuse *correctness* of a model with *soundness* of a planner. Correctness of the model refers to its accuracy in reflecting human perception.

We recruited 78 participants through Mechanical Turk and paid them each $.05 for completing the survey. Because Mechanical Turk is noisy, it is necessary to filter the responses to ensure quality results. We filtered out noise in three ways:

1. For each story, we asked a basic comprehension question, e.g. "How much money did Jessica have after Step 2?", to verify that participants were paying close attention. If they were unable to answer all of these questions correctly, we discarded their data. Those who answered all the comprehension questions correctly were awarded a $0.45 bonus, of which they were made aware from the start.
2. We asked one similarly structured question from a simple propositional domain to verify that the participant understood the type of question we were asking. In this story, a character picks up a key and a sword and goes to a cave where there is a locked treasure chest and a troll. He slays the troll with the sword and opens the chest with the key. Participants were asked why he picked up the key, why he picked up the sword, and why he went to the cave. We expected the participant's answers to reflect the classical definition of causal links—that taking the key is linked to opening the chest, taking the sword is linked to slaying the troll, and going to the cave is linked to *both* opening the chest and slaying the troll. If the participant's answers did not match this, we assumed they either did not understand the type of question we were asking, or they were not paying close attention, and thus we discarded their data.
3. We discarded the data of anyone who answered any question with "So she could buy a car", since a car was never mentioned anywhere in the study and participants had the option of selecting "None of the above". We assume that if they chose the car option, they simply were not paying attention.

Results

After filtering we were left with 20 valid responses to use in our evaluation. We first measured the inter-rater reliability across all questions using Krippendorff's alpha (Krippendorff 2012). While any positive value of *alpha* represents some overall agreement, our result (*alpha* = 0.283) was not strong enough to conclude that there is significant agreement about these questions among humans. This is as we expected, due to the subjective nature of the questions; in general, people do not fully agree on where to draw causal links. We will return to this observation in the Discussion section.

Even in the presence of some disagreement, we were able to identify statistically significant answers for most of the questions. For each question, we used the binomial exact test to determine which, if any, of the four possible answers were more likely to appear. The null hypothesis was that each answer would have a probability of 0.25. Note that it is possible for more than one answer to be significant. In most cases there was exactly one significant answer, but in some cases there were two, and in some cases there were none.

Figures 4-10 show each of the seven stories along with the links drawn by the *significant* answers identified from this test. Each outgoing link or pair of links is labeled with the p-value for that answer. If the significant answer was "none",

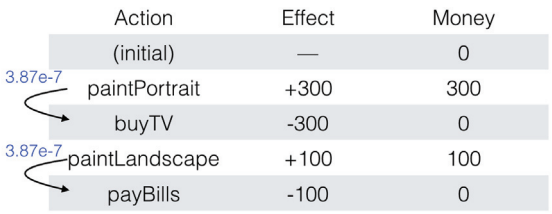

Figure 4: Significant Answers: Question 1

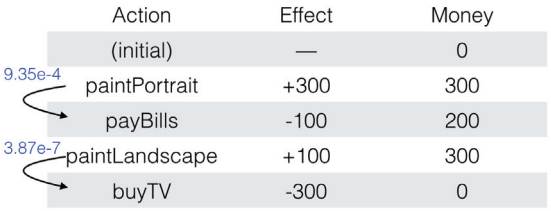

Figure 5: Significant Answers: Question 2

this is represented by a bold **x** on the right side of the step number. The two items in Question 7 which have *both* a link on the left and an **x** on the right are the questions for which there were two significant answers. The items labeled with only a question mark are those for which there were no significant answers.

We consider these the correct answers for our scoring procedure. If a question has two correct answers, a model scores points for answering either of them. We do not score the questions for which there are no correct answers.

Our scoring procedure is as follows. For each question, there are two possible links to be drawn: one for *buyTV* (henceforth shortened to just "tv") and one for *payBills* (henceforth just "bills"). A model can score up to two points for each answer: one point for each correctly drawn link, where "correctly drawn" includes *not* drawing a link which is *not* supposed to be drawn. In other words, if the correct answer is "bills" this means that the *bills* link should be drawn, and the *tv* link should NOT be drawn. A model scores one point for each of these links it handles correctly.

Figure 11 shows the final scores of all ten models, as well as the number of unexplained actions[1] for each model (to be used as a tie-breaker).

Discussion

The two highest scoring models were IMM / ACC and SKP / ACC, which use the same rule for stopping (account for the precondition only) but different rules for starting. We chose to break ties by looking at the number of unexplained actions. For our purposes, we would prefer to have fewer unexplained actions, so we consider the winning model to be SKP / ACC. In this model, we link the most recent action for which the precondition was not already true, and then continue linking until we account for the amount required by the

[1] An unexplained action in this case is any *paint* action for which the model created no outgoing causal links.

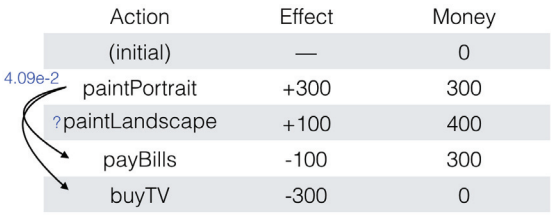

Figure 6: Significant Answers: Question 3

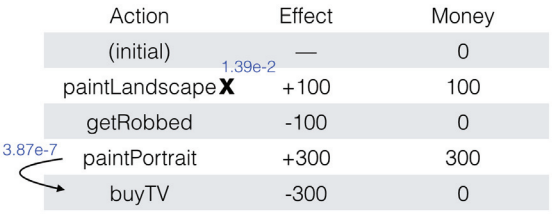

Figure 7: Significant Answers: Question 4

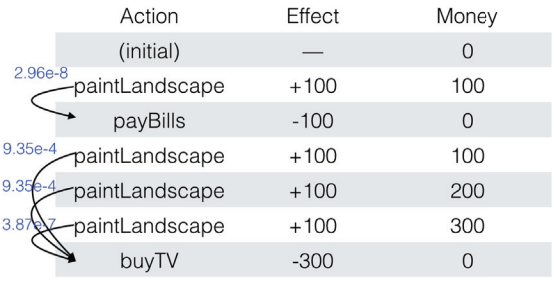

Figure 8: Significant Answers: Question 5

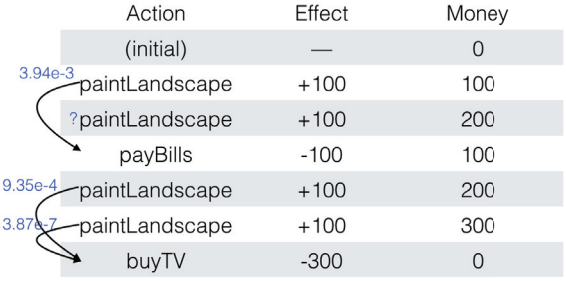

Figure 9: Significant Answers: Question 6

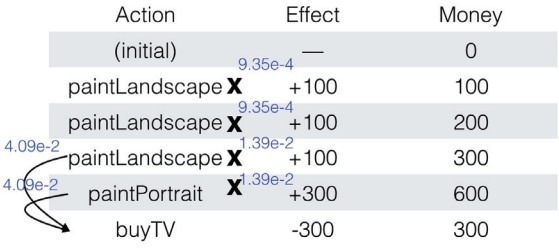

Figure 10: Significant Answers: Question 7

Model	Score / 36	Unexplained Actions
IMM / ONE	30	9
SKP / ONE	31	8
IMM / ACC	33	5
SKP / ACC	33	3
IMM / ACC+	32	4
SKP / ACC+	32	3
IMM / ACC++	29	3
SKP / ACC++	30	2
IMM / ALL	29	0
SKP / ALL	30	2

Figure 11: Models Scored against Correct Answers

precondition. The algorithm for generating this set of causal links is given in Algorithm 1.

Algorithm 1 SKP / ACC

Given a plan and a step s_{head} in that plan with precondition $(v \geq k)$:
Let $k_{needed} = k$
Let $k_{counted} = 0$
Let $s = s_{head}$
while $k_{counted} < k_{needed}$ **do**
    Let s' be the last step prior to s that increased v
    Let v' be the value of v in the state before s'
    if $v' < k$ **then**
        Let k' be the amount by which s' increased v
        Add link $s' \xrightarrow{v} s_{head}$
        $k_{counted}$ += k'
    end if
    Let $s = s'$
end while

However, we learned from our inter-rater reliability coefficient that agreement between individuals on this topic is not particularly strong. Furthermore, most of the models we tested performed fairly well in our analysis. We feel that there is sufficient evidence to say that the SKP / ACC+ model, for example, is reasonably accurate at representing the way humans perceive enablement. We like this model in particular because it accounts for changes to the fluent in the opposite direction. It also creates more links than its SKP / ACC counterpart and is therefore more conservative, leaving us more possible ways to explain character actions.

For example, consider the problem in Figure 9, which consists of four *paintLandscape* actions with a *payBills* action in the middle of them. When generating links for the final *buyTV* action, both of these models link the last three *paintLandscape* actions, but only SKP/ACC+ links the first one as well. It seems a reasonable story that the character paints all four landscapes for the same goal—because she wants a new TV—despite the fact that she must spend some of it on bills in the process. The algorithm for generating links using this model is given in Algorithm 2.

Algorithm 2 SKP / ACC+

Given a plan, and a step s_{head} in that plan with precondition $(v \geq k)$:
Let k_{needed} = k
Let $k_{counted} = 0$
Let $s = s_{head}$
while $k_{counted} < k_{needed}$ **do**
    Let s' be the last step prior to s that modified v
    Let k' be the amount by which s' modified v
    if k' is negative **then**
        k_{needed} += $\mid k' \mid$
    else if k' is positive **then**
        Let v' be the value of v in the state before s'
        if $v' < k$ **then**
            Add link $s' \xrightarrow{v} s_{head}$
            $k_{counted}$ += k'
        end if
    end if
    Let $s = s'$
end while

Limitations

This was a preliminary study and has some limitations that are worth noting. First, we ended up with a very small sample size after filtering. A larger study may achieve stronger agreement between subjects and provide a more accurate human answer set. Second, we did not explore any type of numeric quantities other than money. Humans may treat other types of values differently. Third, we did not explore the idea of debt, and whether humans treat negative numbers or zero as a special case. The ACC* models only draw enough causal links to account for some positive number, meaning they treat zero as a stopping point. Humans might consider it a continuous scale.

Acknowledgements

This research was supported by NSF award IIS-1464127.

References

Cardona-Rivera, R. E.; Cassell, B. A.; Ware, S. G.; and Young, R. M. 2012. Indexter: a computational model of the event-indexing situation model for characterizing narratives. In *Proceedings of the 3rd Workshop on Computational Models of Narrative*, 34–43.

Coles, A. J.; Coles, A. I.; Fox, M.; and Long, D. 2012. Colin: Planning with continuous linear numeric change. *Journal of Artificial Intelligence Research* 44:1–96.

Eyerich, P.; Mattmuller, R.; and Roger, G. 2012. Using the context-enhanced additive heuristic for temporal and numeric planning. In *Towards Service Robots for Everyday Environments*, volume 76 of *Springer Tracts in Advanced Robotics*. Springer. 49–64.

Fox, M., and Long, D. 2003. Pddl2.1: An extension to pddl for expressing temporal planning domains. *J. Artif. Intell. Res. (JAIR)* 20:61–124.

Gerevini, A. E.; Saetti, A.; and Serina, I. 2008. An approach to efficient planning with numerical fluents and multi-criteria plan quality. *Artif. Intell.* 172(8-9):899–944.

Gratch, J. 2000. Émile: Marshalling passions in training and education. In *Proceedings of the Fourth International Conference on Autonomous Agents*, 325–332. ACM.

Helmert, M. 2006. The fast downward planning system. *Journal of Artificial Intelligence Research* 26:191–246.

Krippendorff, K. 2012. *Content analysis: An introduction to its methodology*. Sage.

McAllester, D., and Rosenblitt, D. 1991. Systematic nonlinear planning. Technical report, Massachusetts Institute of Technology Artificial Intelligence Laboratory.

Riedl, M. O., and Young, R. M. 2010. Narrative planning: balancing plot and character. *Journal of Artificial Intelligence Research* 39(1):217–268.

Trabasso, T., and Sperry, L. L. 1985. Causal relatedness and importance of story events. *Journal of Memory and language* 24(5):595–611.

Trabasso, T., and Van Den Broek, P. 1985. Causal thinking and the representation of narrative events. *Journal of memory and language* 24(5):612–630.

Ware, S. G., and Young, R. M. 2014. Glaive: a state-space narrative planner supporting intentionality and conflict. In *Proceedings of the 10th AAAI International Conference on Artificial Intelligence and Interactive Digital Entertainment*, 80–86.

Ware, S. G.; Young, R. M.; Harrison, B.; and Roberts, D. L. 2014. A computational model of narrative conflict at the fabula level. *IEEE Transactions on Computational Intelligence and Artificial Intelligence in Games* 6(3):271–288.

Young, R. M.; Ware, S. G.; Cassell, B. A.; and Robertson, J. 2013. Plans and planning in narrative generation: a review of plan-based approaches to the generation of story, discourse and interactivity in narratives. *Sprache und Datenverarbeitung, Special Issue on Formal and Computational Models of Narrative* 37(1-2):41–64.

Young, R. M. 1999. Notes on the use of plan structures in the creation of interactive plot. In *Proceedings of the AAAI Fall Symposium on Narrative Intelligence*, 164–167.