Validating a Plan-Based Model of Narrative Conflict

Stephen G. Ware
Liquid Narrative Research Group
North Carolina State University
Raleigh, North Carolina, USA
sgware@ncsu.edu

R. Michael Young
Liquid Narrative Research Group
North Carolina State University
Raleigh, North Carolina, USA
young@csc.ncsu.edu

ABSTRACT

Conflict is an essential element of interesting stories. In previous work, we proposed a formal model of narrative conflict based on intentional planning that is designed to facilitate story generation. This paper presents the results of an experiment to test whether or not our model can answer "who?" "why?" and "when?" questions about conflict similarly to humans analyzing the same stories. Our model has some success at predicting which conflicts human readers will report and performs well at recognizing which characters are in conflict and for how long.

Categories and Subject Descriptors

I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods

General Terms

Conflict Modeling, Computational Narratology

Keywords

Conflict, Narrative, Participants, Subject, Duration

1. INTRODUCTION

Narratologists, screen writers, game designers, and other researchers in computer narrative agree that conflict is an essential element of interesting stories [15, 5, 17, 9, 2, 1]. Conflict provides an impetus for the plot to begin [6], and it keeps the audience engaged as the story unfolds, even if they already know the ending [8]. Conflict also structures the discourse of a story into meaningful units that together make up a coherent whole [6, 1].

Narrative-oriented virtual environments like RPG games, training simulations, and interactive tutoring systems often need to adapt their stories to respond to users. A built-in understanding of the story structure provided by conflict can benefit these systems by allowing them to modify their content so that it is still well-structured and engaging for the audience.

Previously, we defined a formal model of narrative conflict called CPOCL (Conflict Partial Order Causal Link) which was inspired by research in narratology and based on AI planning [20]. We also presented an algorithm that makes use of this model to generate stories with conflicts.

In this approach, conflict occurs when a goal seeking agent's plan is thwarted by another agent's plan, the environment, or its own plans to achieve other goals. CPOCL is able to answer essential questions about the conflicts it models. It can identify which characters are involved, why their plans are incompatible, when a conflict begins, and when a conflict gets resolved. In order to demonstrate that CPOCL accurately reflects a human understanding of this important narrative phenomenon, we designed an experiment in which participants read three different fictional short stories and reported details about all the conflicts they noticed. Their answers were compared to the conflicts in those stories that were modeled by CPOCL.

The model performs well enough to support a link between CPOCL structures and the human perception of conflict. Where it fails, the data offers insights about how the model should be extended with work from non-classical planning and psychology.

2. RELATED WORK

As far back as Talespin [11] and as recently as PaSSAGE [18], narrative generation systems have relied on human authors to supply the conflict that drives the story. This is also common in the games industry; most story-based games have a pre-established plot. While stories may branch based on user choices, the content of these branches are usually also pre-established at design time.

This method has been generalized by systems like Universe [10] and Mexica [12], which combine pre-scripted plot fragments (or plot grammars) to produce whole stories. However, the general problem of building well-structured plot fragments from atomic actions remains unsolved. Systems which rely on pre-scripted plots or plot fragments rely on their authors to model conflict implicitly. By making conflict explicit in the model, we gain a greater ability to reason about this essential phenomenon and adapt interactive stories.

Smith [16] generated conflict by using an adversarial game-playing algorithm to create the story. This approach oversimplifies the antagonist; it is not simply a malevolent force to make trouble for the protagonist. The antagonist

FDG '12 May 29-June 1, 2012 Raleigh, NC, USA.

has its own goals, and should thwart the protagonist only when those goals require it to do so.

Zambetta, Nash, and Smith [24] model conflict in stories as a system of differential equations that simulate an arms race scenario. This approach is helpful as high-level control for the pace of a story, but it cannot reason about the individual motivations of the participants. Similarly, the AI director in Valve's *Left 4 Dead* series [3] adapts the number of enemies, frequency of health kits, and even the layout of the level to suit individual players. The intensity of conflict can be increased and decreased, but the fundamental story of survivors vs. zombies remains the same.

Barber and Kudenko [2] create dramatic tension in their GADIN system with *dilemmas*—decisions the user must make which will negatively affect at least one character. GADIN detects when these dilemmas are applicable to the story and applies them to engage the user. Similarly, Szilas [17] annotates actions in his IDtension system with a "conflict value" that measures to what extent the action forces a character to violate its moral values.

These methods represent progress toward encoding an understanding of conflict into the story generation process. However, dilemmas are a subset of all the conflicts available to story writers. They arise and get resolved immediately, so it is difficult to model the thematic and extended conflicts that provide important macro-structural features of a story.

3. THE CPOCL MODEL OF CONFLICT

The *fabula* of a story is a complete set of events with their causal and temporal constraints [9]. Partial Order Causal Link (POCL) plans have been a popular data structure for expressing stories because they explicitly represent this information in a formal model that can be generated and manipulated by machines [23].

Riedl and Young [13] extended POCL planning to model the intentional behavior of characters. Ware and Young [20] further extended that work to model conflict. This section provides a brief overview of the CPOCL model (see [20] for full details).

3.1 Threatened Causal Links as Conflict

Stories are composed of STRIPS-style **steps** [7]. Each step has a set of preconditions which must be true before the step can occur and a set of effects that represent how the world will change after the step occurs. Preconditions and effects are described in the language of function-free first order ground predicate literals. Each story has one placeholder **start step**, which has only effects that express the initial state, and one placeholder **end step**, which has only preconditions that must be true by the end [22].

Logical constants represent objects in the story. Some subset of those constants represents the story's intentional agents, or **characters**.

An **intention frame** is a set of steps that some character c intends to carry out in order to achieve one of its personal goals g [13]. All steps in the story which require the consent of a character must appear in an intention frame for that character. In other words, each action that a character takes must be in service of one of its personal goals. An intention frame has a **motivating step** with the effect $intends(c, g)$ that explains why the character adopted the goal. An intention frame also has a **satisfying step** with effect g that explains how the character achieved the goal.

Executed steps actually occur—that is, they will be narrated as part of the story. **Non-executed steps** are intended by some character but never occur because one or more of their preconditions are not satisfied. Though non-executed steps do not occur, they are an important source of information about the inner lives of characters and can represent alternate story paths.

A **causal link** $s \xrightarrow{p} t$ joins two steps s and t. Step s must have the effect p, and step t must have the precondition p, and s must occur before t. Causal links explain how the preconditions of a step get satisfied by the earlier events of the story. The literal established by a causal link (in this case, p) is called its **label**.

A causal link becomes a **threatened causal link** when a third step u is ordered after step s but before step t and has an effect $\neg p$. To say that a causal link is threatened is to say that the literal it establishes gets undone before it is needed in the story.

Traditional planners must remove all threatened causal links from a story in order to guarantee soundness. By marking some steps in a plan as intended but non-executed, the CPOCL algorithm is able to retain some threatened causal links while still ensuring that the story is causally and temporally sound [20].

Given these terms, we say that a **conflict** occurs just when:

- A character c_1 intends a step t as part of intention frame f_1 whose motivating step m_1 is executed.
- A character c_2 intends a step u as part of intention frame f_2 whose motivating step m_2 is executed, and $f_2 \neq f_1$.
- There exists some causal link $s \xrightarrow{p} t$ threatened by u.
- Either t or u or both are non-executed steps.

Simply put, a conflict occurs when one character's plan prevents (or attempts to prevent and fails) another character's plan from succeeding. Since the two plans are incompatible, at most one can succeed. It is also possible that $c_1 = c_2$, which is often called *internal conflict*.

Some steps are not intended by anyone, such as accidents and the forces of nature. For convenience, we group all these steps into a single intention frame which is intended by Fate (who becomes a personified character). Actions intended by Fate that conflict with a character's plans cause conflict with the environment or with destiny.

3.2 Example: Stealing the Antivenom

An example conflict from the Western story used in the experiment (see Figure 2) will help to illustrate the CPOCL model. In the first step of the story, Hank's young son Timmy gets bitten by a snake and becomes sick. This step becomes the motivating step for one of Hank's plans; the satisfying step occurs when Hank heals his son. In order to accomplish this, Hank intends to steal some antivenom from the local general store owner, Carl.

Hank's theft motivates William, the town sheriff, to arrest Hank, take back the antivenom, and return the stolen property to Carl. Excerpts from Hank's plan and William's plan are illustrated in Figure 1. One effect of Hank's "Steal Antivenom" action is that Hank has the antivenom. He needs the antivenom in order to heal his son, so a causal link is drawn from the "Steal Antivenom" step to the "Heal

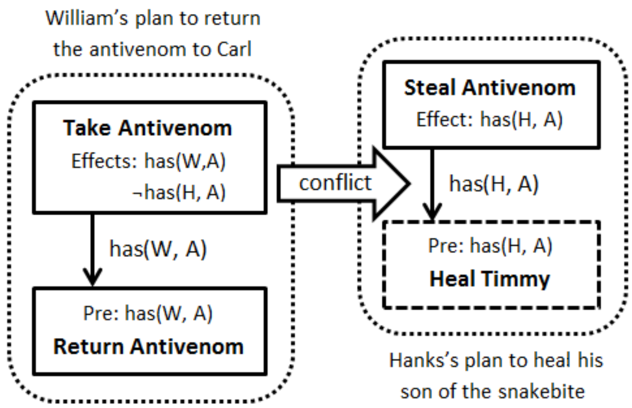

Figure 1: An example conflict taken from the Western story. Sheriff William's (W) plan to return the antivenom (A) to Carl thwarts Hank's (H) plan to heal his son, Timmy.

Timmy" step. However, William's "Take Antivenom" step threatens this causal link, because if William takes the antivenom Hank will no longer have it.

"Heal Timmy" is a non-executed step (indicated by the dashed border), meaning that it does not occur despite Hank's efforts. The threat to the causal link in Hank's plan is a conflict, and since Hank's plan eventually fails, it does not violate the temporal and causal consistency of the story as a whole.

3.3 Maintenance Goals

One extension was made to CPOCL for this experiment in order to represent maintenance goals—literals which are currently true that a character wishes to remain true. A **persistence step** has exactly one precondition and one effect that are identical. All persistence steps occur simultaneously with the story's end step.

For example, Timmy is alive at the beginning of the Western story and wishes to remain alive until the end. Formally, the start step has the effects:

$alive(Timmy) \wedge intends(Timmy, alive(Timmy))$

Since this goal is already satisfied, Timmy does not need to carry out a plan to make it true. However, an intention frame is still needed to expresses Timmy's desire, so a persistence step with the precondition and effect $alive(Timmy)$ is used as the satisfying step of Timmy's intention frame. A causal link can be drawn from the start step of the story to the precondition of the persistence step. This way, any step with the effect $\neg alive(Timmy)$ will come into conflict with Timmy's maintenance goal to stay alive.

3.4 Participants, Subject, and Duration

CPOCL's broad definition of conflict is intended to cover the spectrum of all narrative conflicts. Ware and Young [19] also compiled seven dimensions that distinguish one conflict from another. The first three—*participants*, *subject*, and *duration*—are structural properties of the model. They answer "who?" "why?" and "when?" respectively. This paper presents an evaluation of these three dimensions to demonstrate that threatened causal links in CPOCL plans correspond to the human perception of conflict in stories.

The last four dimensions—*balance*, *directness*, *intensity*, and *resolution*—are quantitative and require additional context information to measure. Some initial evaluation has already been conducted for these dimensions [21].

Participants.

The *participants* of a conflict, labeled c_1 and c_2, are the two characters associated with the conflicting intention frames. Recall that for internal conflicts, $c_1 = c_2$.

Subject.

The *subject* of a conflict is the condition which makes the two character plans incompatible—the label of the threatened causal link. Textually, it can be expressed as "c_1 intends step t, which requires p, but c_2 intends step u, which causes $\neg p$." For example, "Hank intends to heal Timmy, which requires him to have the antivenom, but William intends to take the antivenom from Hank, which causes Hank not to have the antivenom."

One important direction of future work will be to reason about the subject of conflict at a higher level. For example, a reader might say that William's duty to uphold the law is in conflict with Hank's duty to care for his son, or they might say that it is a conflict between the letter of the law and the spirit of the law. CPOCL is not yet able to reason about the subject of a conflict at this level of abstraction.

Duration.

The *duration* of a conflict is the interval of time during which both participants intend their incompatible plans. The steps of a CPOCL plan are partially ordered, so to calculate duration, some valid total ordering O is nondeterministically chosen. Let index(s, O) be the index of step $s \in O$ such that the placeholder start step has index 0, the first step has index 1, the second step index 2, and so on until the placeholder end step, which has index n. By definition, all persistence steps also have index n.

A story can now be envisioned as a sequence of n states. t_0 is the initial state of the story, occurring before the first step (i.e. the step with index 1). t_1 is the state after step 1 has occurred, t_2 the state after step 2, etc. The duration of a conflict is the number of states during which c_1 intends f_1 and c_2 intends f_2. To determine this, we need to know when intention frames begin and end. The beginning is simply the state after the motivating step, but detecting the end is more complicated.

The end of an intention frame is the state by which a character has abandoned its plan. If all of the steps in an intention frame are executed, the frame ends once the last step is executed. If some of the steps in the frame are non-executed, the frame ends after the last executed step. One important exception to this rule exists: if the first non-executed step in a frame is step t of a conflict (the head step of a threatened casual link), then the intention frame ends after step u (the threatening step). The reason for this exception comes from the nature of conflict: if a character abandoned a plan because it was thwarted, he should intend the plan up until the time when the plan gets thwarted.

Let the function $\Omega(f)$ return the index of the state by which intention frame f has ended. Recall from section 3.1 that m_1 and m_2 are the motivating steps of the two conflicting intention frames. Now, we can define the duration of a conflict as:

$$\begin{aligned}\text{start} &= \max(\text{index}(m_1), \text{index}(m_2))\\ \text{end} &= \min(\text{index}(t), \text{index}(u), \Omega(f_1), \Omega(f_2))\\ \text{duration} &= \text{end} - \text{start}\end{aligned}$$

An example will help to make this more clear. When young Timmy gets bitten by a snake in the first step of the Western story, Hank forms a plan to heal him by stealing some antivenom. The state after the first step is t_1. Everything goes well for Hank until he steals the antivenom from Carl in the fourth step. This theft causes sheriff William to form a plan to return the antivenom to Carl. Now, at time t_4, Hank's intention frame comes into conflict with William's.

This conflict gets resolved when William wins out over Hank by tying him up in step 6. Even though Hank intends steps with an index higher than 6, such as the non-executed step where he heals Timmy, he is forced to abandon his plan at time 6 when he is subdued by William.

4. EVALUATION

We designed an experiment in which human users[1] analyzed the conflicts in three short, fictional stories. One takes place in an American western setting (Western), one in a medieval fairy tale kingdom (Fantasy), and one in a futuristic science fiction setting (Space). Readers analyzed conflicts in terms of "who?" "why?" and "when?" questions. We used this data to evaluate two hypotheses:

1. Users will report conflicts similarly to one another.
2. Users will report conflicts that are similar to those defined by CPOCL.

4.1 Experimental Design

The experiment was conducted as an online survey, and users were recruited via e-mail and social networking websites. No compensation or incentives were offered. Users had to complete a tutorial in order to familiarize themselves with the interface. After that, each user reported conflicts for all three stories, which were presented in a random order.

4.1.1 Interface

The interface for reporting conflicts is shown in Figure 3. It has three main components. The top left box displays the story up until the current time. Users begin at t_0 and can move backward and forward through time (i.e. through world states). The bottom left box contains images of all the characters involved in the story (including Fate). Each character has a thought bubble that contains his or her plan at that moment. Users were given access to this information in order to make the *fabula* of the story as clear as possible. CPOCL is a model of the *fabula*, and we wished to avoid any confusion that might arise if users made different predictions about what a character knows or plans to do.

The right box is a list of conflicts. When a user notices a conflict in the story, he clicks the "Add New Conflict" button and is then prompted for four things:

- A first character
- A second character

[1]In order to avoid confusion with the *participants* of a conflict, we refer to those people who participated in our experiment as "users."

Western Story
Once upon a time in the Wild West, there lived a cattle rancher named Hank and his young son Timmy. Not far from their ranch was a small town that had a saloon and a general store. William was the sheriff of the town, and it was his job to arrest and imprison anyone who broke the law. Carl owned the general store, and he sold all sort of things, including a powerful antivenom to cure snakebites. Then, one day...

1. Hank's son Timmy got bitten by a snake and became sick.
2. Hank went to the general store.
3. Hank tied up Carl.
4. Hank stole the antivenom from Carl.
5. William went to the general store.
6. William tied up Hank.
7. William took the antivenom from Hank.
8. William untied Carl.
9. William gave the antivenom to Carl.
10. William took Hank to jail.
11. Timmy died of his snakebite.

The end.

Fantasy Story
Once upon a time in a small village there lived a beautiful maiden named Talia. She was in love with a handsome thief named Rory, but Rory was too poor to support her. One day, Talia caught the eye of the kingdom's prince, Vince, and he also fell in love with Talia. Talia did not love Prince Vince, but he was very rich. Then one day...

1. Rory proposed to Talia.
2. Prince Vince proposed to Talia.
3. Gargax got hungry.
4. Gargax went to the village.
5. Gargax devoured Prince Vince.
6. Rory went to Gargax's cave.
7. Rory stole Gargax's treasure.
8. Rory went to the village.
9. Talia married Rory.

The end.

Space Story
Many years in the future, space explorers will travel from planet to planet attempting to make peaceful contact with alien races. This is the story of Zoe, a space explorer orbiting the planet Mydrox in her starship. Deep in a cave on the planet Mydrox lives a dangerous Lizard Beast. One day...

1. Zoe teleported to the surface of planet Mydrox.
2. The Lizard Beast walked to the surface of planet Mydrox.
3. The Lizard Beast started a fight with Zoe, which made Zoe angry.
4. Zoe calmed the Lizard Beast with a soothing song.
5. A massive volcano on the surface of planet Mydrox began to erupt.
6. Zoe teleported to her ship.
7. The Lizard Beast walked to its underground nest.
8. A massive volcano erupted, covering the surface of planet Mydrox with magma, but no one was killed.

The end.

Figure 2: The three stories read by users in the experiment. Due to space constraints, each individual character's plans are not displayed.

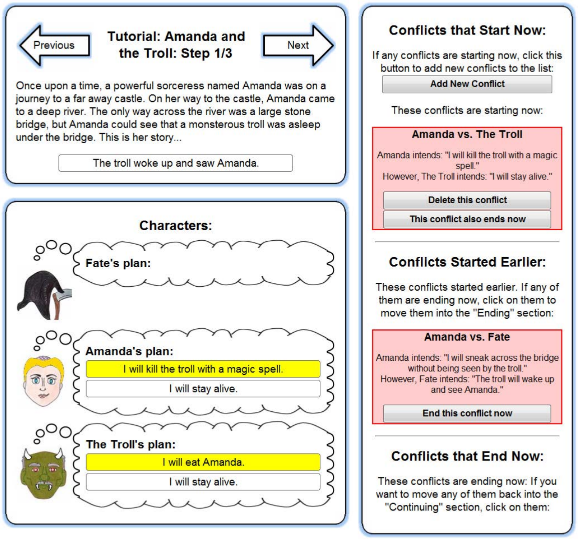

Figure 3: The interface by which users read stories and reported conflicts.

- An action from the first character's thought bubble
- An action from the second character's thought bubble that thwarts the first character's plan

This information provides the *participants* and *subject*[2] of the conflict.

The list of conflicts is separated into three parts: those starting at this moment in the story, those that started earlier, and those ending at this moment. New conflicts appear in the "start now" category. When the user moves forward in time, those conflicts appear in the "started earlier" category. At every moment, each conflict has a button to "end this conflict now," which moves it into the "end now" category. This allows the user to visually report the *duration* of a conflict.

To summarize, for each story each user reports any number of conflicts. A conflict is a 6-tuple $\langle c_1, c_2, s_1, s_2, b, e\rangle$, where c_1 is a character who plans step s_1, c_2 is a character who plans step s_2, b is index of the state in which the conflict begins, and e is the index of the state by which the conflict has ended. The order of participants is not important—in other words $\langle c_1, c_2, s_1, s_2, b, e\rangle = \langle c_2, c_1, s_2, s_1, b, e\rangle$.

4.1.2 Response

23 users responded to the survey by finishing one or more stories. Of those, 19 users finished all three stories. There were 13 male users and 10 female users with a median age range of 26-35. In total, 408 conflicts were reported across the three stories. If a user reported no conflicts for a story, that user's data was not included in the analysis for that story.

[2]Users may report any two steps that meet these criteria. This makes it possible to report two steps which do not thwart one another—that is, no effects of the first step negate any preconditions of the second step and vice versa. However, none of these so-called invalid conflicts were ever reported by enough users to be considered correct according to Section 4.3.

Table 1: Fleiss's κ (user agreement) for three stories.

		exact		overlap	
Story	**users**	**# of ?s**	κ	**# of ?s**	κ
Western	21	63	0.26	31	0.45
Fantasy	20	31	0.32	18	0.61
Space	22	48	0.15	21	0.43
Average			**0.24**		**0.50**

4.2 Inter-User Agreement

Before evaluating CPOCL based on user data, it is important to establish that users agree amongst themselves about the conflicts they reported. For this we used Fleiss's κ coefficient, which measures agreement among multiple users. κ reaches 1 if users agreed completely and reaches -1 if users disagreed completely. Fleiss's κ assumes that three or more people are answering some number of multiple-choice questions, so we developed a way to interpret our data in this fashion.

The most straightforward interpretation would be to consider every conflict that could possibly have been reported as a question that was implicitly answered as *true* if the user reported it or *false* if the user did not report it. However, this would artificially inflate the κ value with an abundance of true negatives (conflicts which were possible to report but were not reported). To account for this, we only considered those conflicts which were reported by at least 1 user as the range of all possible conflicts. This was less than 1% of all the conflicts that could possibly have been reported.

The first column of Table 1, labeled *exact*, shows the κ values achieved for each story. *Users* is the number of users who finished that story. *# of ?'s* is the number of possible conflicts (i.e. number of questions) to which users implicitly answered *true* or *false*.

Many of the conflicts reported by users had the same participants, same subject, and overlapping (but not exactly the same) duration. In order to account for this, we calculated a second set of κ values such that these conflicts were considered the same. The results are shown in the second column of Table 1, labeled *overlap*. Allowing for overlapping duration reduced the range of reported conflicts by about half for each story and significantly increased the κ values.

Based on these results, we conclude that users demonstrated some agreement about which conflicts exist in the three stories, especially when allowing for overlapping durations.

4.3 User Agreement with CPOCL

We define a conflict to exist according to human readers if it was reported by at least one third of users (overlapping durations considered the same). This set of conflicts for each story was the standard against which we tested CPOCL's performance on two tasks.

To our knowledge, this is the first formal model of plan-based conflict. Since there are no similar approaches against which to compare CPOCL, we compare it to both a naïve baseline and the performance of the average individual user.

4.3.1 Prediction

Table 2: Confusion matrices for CPOCL's predictions compared to human readers.

Western		Fantasy		Space	
TP: 10	FP: 4	TP: 4	FP: 4	TP: 6	FP: 3
FN: 0	TN: 17	FN: 0	TN: 10	FN: 0	TN: 12

Table 3: CPOCL's accuracy (Acc.), precision (Pre.) and recall (Rec.) for the prediction and recognition tasks.

	Prediction			Recognition		
Story	Acc.	Pre.	Rec.	Acc.	Pre.	Rec.
Western	0.87	0.71	1.00	0.98	0.90	1.00
Fantasy	0.78	0.50	1.00	0.90	0.61	1.00
Space	0.86	0.67	1.00	0.87	0.82	0.86
Average	**0.84**	**0.63**	**1.00**	**0.92**	**0.77**	**0.95**

For the first task, which we call *prediction*, we treated CPOCL as a user and compared the set of conflicts that it defines to those reported by users. The resulting confusion matrices are shown in Table 2. A true positive is a conflict defined by CPOCL that one third of users reported. A false positive is a conflict defined by CPOCL that one third of users did not report. A false negative is a conflict reported by one third of users that CPOCL does not define. A true negative is a conflict which was reported by at least one user but less than one third of users and which CPOCL does not define. Summary statistics for the prediction task are presented in Table 3.

CPOCL performs relatively well on this task considering the extremely low probability of guessing correctly. We define a random guess as follows: Choose two characters from the story at random; choose a start and end time at random such that the start time is less than or equal to the end time; choose two actions at random such that the first action is from one of the first character's intention frames, the second action is from one of the second character's intention frames, and both actions occur after the start time. For all three stories, the chances that a random guess was correct according to one third of users was less than 0.02% (about 1 in 5000), even allowing for overlapping durations.

We can also compare the model's performance to that of each individual user. Precision (true positive rate) is the most meaningful statistic for this comparison because it expresses the fraction of conflicts reported that were correct. These results are visualized in Figure 4. In short, CPOCL does significantly better than random guessing (a very naïve baseline) but not quite as well as the average human reader (the ideal).

4.3.2 Recognition

Another way to evaluate the model is to test how well it can recognize when a given pair of characters are in conflict. This task, called *recognition*, asks this question both of users and of CPOCL: for every state, and for every pair of characters, are those characters in conflict in that state?

Figure 5 presents a visualization of the results. True positives indicate that users and CPOCL both answered

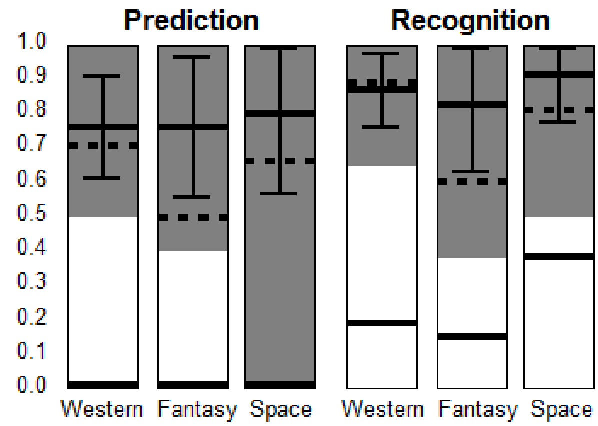

Figure 4: CPOCL's precision vs. a naïve baseline and individual human users. The gray region represents the range of precision values for individual users. The higher solid black bar is the average user precision (+/- one standard deviation). The lower solid black bar is the precision of the baseline. The dashed black bar is CPOCL's precision.

"yes." True negatives indicate that users and CPOCL both answered "no." False positives indicate that CPOCL answers "yes," but users answers "no." False negatives indicate that users answers "yes," but CPOCL answered "no." Summary statistics for the recognition task are presented in Table 3.

A naïve baseline for this task is to always answer "yes" or "no" to every question. Of those two, answering "no" will yield the highest accuracy, and answering "yes" will yield the highest precision. We compared CPOCL's performance to these two models, and the results are presented in Table 4.

When compared to individual human users, CPOCL did better on the recognition task and even outperformed the average reader for the Western story (see Figure 4).

4.4 Discussion

4.4.1 Limitations of Classical Planning

Some conflicts defined by the model are understandably counter-intuitive to users due to a mismatch between how people think about actions and the knowledge representation of a STRIPS-style story domain. For example, CPOCL defines the following conflict in the Western story: William intends to take the antivenom from Hank, but Hank intends to travel back to his ranch. At first glance, there does not seem to be any conflict here.

One precondition for the "take" action is that both the person who has the item and the person who is taking the item be at the same location (in this case, the general store). If Hank travels back to his ranch, he will no longer be at the general store and the "take" action will fail. William can still take the antivenom from Hank, but he can no longer take the antivenom from Hank *at the general store*.

This suggests that users do not think about steps in terms of their exact mechanics; rather, they think at a more abstract level. Going to the general store from the ranch and going to the general store from the saloon can

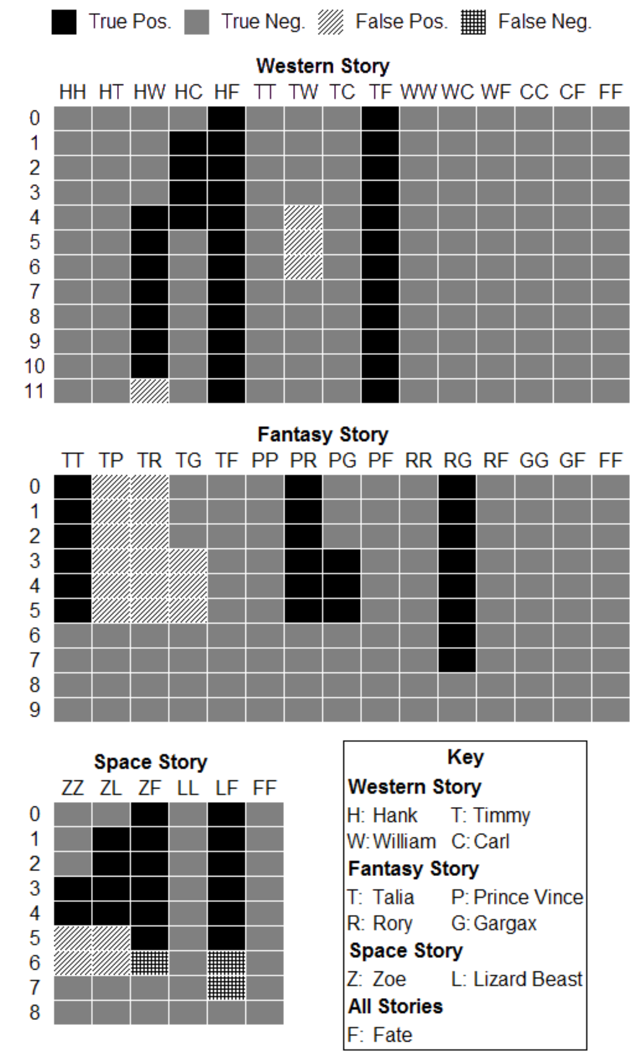

Figure 5: A visualization of CPOCL's performance on the recognition task. For each story, the vertical axis is labeled with the index of the state. The horizontal axis is every pair of characters.

Table 4: CPOCL's (CP) performance on the recognition task relative to two naïve baselines: always answer "no," and always answer "yes." This table shows CPOCL's percent improvement over those baselines.

	Accuracy			Precision		
Story	**No**	**CP**	**%Imp**	**Yes**	**CP**	**%Imp**
Western	0.81	0.98	21%	0.19	0.90	462%
Fantasy	0.85	0.90	6%	0.15	0.61	395%
Space	0.61	0.87	42%	0.39	0.82	110%
Average			**23%**			**322%**

both be abstracted as going to the general store—the initial location is unimportant. This insight may help to inform story domain designers.

Another potential source of confusion is that the classical planning model on which CPOCL is built does not support durative actions—that is, all steps are assumed to happen immediately. A durative action must be represented as multiple actions; e.g. Timmy is bitten by a snake and then later dies from the snakebite, or the volcano begins to erupt and then later it erupts.

In the Space story, both Zoe and The Lizard Beast intend to stay safe from natural disasters at their locations. The step where the volcano begins to erupt has the effect that no one in the area is safe. According to CPOCL, this is the *end* of the conflict; the volcano "won" because both Zoe and the Lizard Beast failed to stay safe (which causes them to form new plans to go to safe locations). Users recognized these conflicts between Zoe and Fate and between The Lizard Beast and Fate, however they reported the end of the conflict as the time when each character had reached a new safe location. This seems like the most natural interpretation of the story, so CPOCL may need to be extended with research from automated scheduling [14] to represent durative actions.

4.4.2 A Superset of Conflicts

Interestingly, no false negatives were predicted for any story. Even for the recognition task, the only false negatives that arose were due to the above disagreement about how long the conflict with the volcano should last, not about who was involved. In other words, the conflicts defined by CPOCL are a strict superset of the correct conflicts according to users.

Future work will need to identify a means of filtering the conflicts defined by CPOCL so as to report only those which users will recognize. To do this, we need to discover why users report some conflicts but not others. Certain threatened causal links are very obvious to readers, while others (that are not formally or structurally different) seem not to be obvious at all. We propose three potential explanations.

User Exhaustion.

Many users reported that the survey was mentally exhausting. This may have caused them to report fewer conflicts. Consider the Fantasy story as an example. Talia's plan to marry Rory thwarts Talia's plan to marry Vince. 18 of 20 users reported this internal conflict. However, only 4 users reported that Talia's plan to marry Rory thwarts Vince's plan to marry Talia, despite the fact that the conflict would exist between the *exact same steps*. Similarly, only 4 users reported that Talia's plan to marry Vince thwarts Rory's plan to marry Talia. If these conflicts had been reported, CPOCL's precision on the recognition task would have increased from 0.61 to 0.92 for that story.

We propose a second experiment in which users are shown a list of potential conflicts to mark as "a conflict" or "not a conflict." We suspect that if users are less overwhelmed by the data collection instrument they will report conflicts more consistently and CPOCL will perform better.

One Character Per Action.

In the Western story, 15 of 21 users reported that Hank's plan to heal Timmy is thwarted by William's plan to take back the antivenom. However, only 2 users reported that Timmy's plan to be healed is thwarted by William's plan to take back the antivenom. Both Hank and Timmy intend the "heal" step, and William thwarts both plans, but users did not report a conflict between William and Timmy.

This example, along with Talia's example above, suggests that perhaps users only associate one character with each action. Consider the conflict $\langle c_1, c_2, s_1, s_2, b, e\rangle$ such that step s_2 is intended by multiple characters (like the marriage steps from the Fantasy story, which require the consent of both the bride and groom). Just because character c_1 is in conflict with character c_2 does not imply that c_1 is also in conflict with the other characters who intend step s_2.

Reading Process.

Another explanation for why users recognize some conflicts but not others may lie in the process of reading. Zwaan and Radvansky [25] have developed a psychological model of how users incorporate information as they read a narrative. Their analysis is based on five dimensions: space, causation, intentionality, protagonist, and time.

Consider the Western conflict described above. Hank is the protagonist of his story. He and William are spatially closer, and their conflict happens sooner than the one between William and Timmy. William intends to take the antivenom from Hank because it is his duty as sheriff, but it is not clear that William has any intention of hurting Timmy. Hurting Timmy may be an unintended side effect. The conflict between William and Timmy, while nearly identical in structure to the one between William and Hank, is vastly different in terms of these five dimensions. This may account for why users reported one conflict but not the other. A study of how these dimensions affect the reading process is already underway for plan-based stories [4].

5. CONCLUSION

The results of this experiment partially confirm our hypotheses. Human readers agree about which conflicts exist in stories, and they report conflicts similar to the ones defined by CPOCL. In general, CPOCL performs much better than a naïve baseline, but not as well as the average human reader. We suspect that by addressing the limitations described above, CPOCL can more closely model a human understanding of conflict.

In conclusion, threatened causal links in intentional planning serve as a good basis for modeling the essential narrative phenomenon of conflict. The CPOCL algorithm, which generates stories based on this model, represents progress toward the goal of empowering computer systems to automatically create and adapt plots based on the appealing structural properties that conflict provides.

Previous work has yielded promising results for selecting conflicts based on the quantitative dimensions of *balance*, *directness*, *intensity*, and *resolution* [21]. This will provide even greater control over story content and bring us closer to the day when intelligent narrative systems can effectively adapt their stories to individual audiences.

6. ACKNOWLEDGMENTS

This research was supported by NSF award IIS-0915598.

7. REFERENCES

[1] H. Abbott. *The Cambridge introduction to narrative.* Cambridge U., 2008.

[2] H. Barber and D. Kudenko. Dynamic generation of dilemma-based interactive narratives. In *AIIDE*, 2007.

[3] M. Booth. The ai systems of left 4 dead. In *Keynote, AIIDE*, 2009.

[4] R. Cardona-Rivera, B. Cassell, S. Ware, and R. Young. A computational model of the event-indexing situation model for characterizing narratives. In *Comp. Models of Narrative Workshop*, 2012.

[5] C. Crawford. *Chris Crawford on game design.* New Riders, 2003.

[6] L. Egri. *The art of dramatic writing.* Wildside, 1988.

[7] R. Fikes and N. Nilsson. STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial intelligence*, 2(3/4):189–208, 1971.

[8] R. Gerrig. *Experiencing narrative worlds: On the psychological activities of reading.* Yale U. Pr., 1993.

[9] D. Herman, M. Jahn, and M. Ryan. Conflict. In *Routledge encyclopedia of narrative theory.* Routledge, 2005.

[10] M. Lebowitz. Story-telling as planning and learning. *Poetics*, 14(6), 1985.

[11] J. Meehan. Tale-spin, an interactive program that writes stories. In *IJCAI*, 1977.

[12] R. Pérez and M. Sharples. MEXICA: A computer model of a cognitive account of creative writing. *J. of Experimental & Theoretical AI*, 13(2):119–139, 2001.

[13] M. Riedl and R. Young. Narrative planning: balancing plot and character. *JAIR*, 39:217–268, 2010.

[14] S. Russell and P. Norvig. *Artificial intelligence: a modern approach.* Pearson, 2003.

[15] M. Ryan. *Possible worlds, artificial intelligence, and narrative theory.* Indiana U., 1991.

[16] T. Smith and I. Witten. A planning mechanism for generating story text. *Literary and Linguistic Computing*, 2(2):119–126, 1987.

[17] N. Szilas. IDtension: a narrative engine for Interactive Drama. In *TIDSE*, 2003.

[18] D. Thue, V. Bulitko, and M. Spetch. Making stories player-specific: Delayed authoring in interactive storytelling. In *ICIDS*, pages 230–241, 2008.

[19] S. Ware and R. Young. Modeling Narrative Conflict to Generate Interesting Stories. In *AIIDE*, 2010.

[20] S. Ware and R. Young. CPOCL: A Narrative Planner Supporting Conflict. In *AIIDE*, 2011.

[21] S. Ware and R. Young. Initial Results for Measuring Four Dimensions of Narrative Conflict. In *INT*, 2011.

[22] D. Weld. An introduction to least commitment planning. *AI magazine*, 15(4):27–61, 1994.

[23] R. Young. Notes on the use of plan structures in the creation of interactive plot. In *AAAI Fall Symp. on Narrative Intelligence*, pages 164–167, 1999.

[24] F. Zambetta, A. Nash, and P. Smith. Two families: dynamical policy models in interactive storytelling. In *ACIE*, pages 1–8. RMIT U., 2007.

[25] R. Zwaan and G. Radvansky. Situation models in language comprehension and memory. *Psychological bulletin*, 123(2):162, 1998.