A Computational Model of Plan-Based Narrative Conflict at the Fabula Level

Stephen G. Ware, *Member, IEEE*, R. Michael Young, *Senior Member, IEEE*, Brent Harrison, *Student Member, IEEE*, and David L. Roberts

***Abstract*—Conflict is an essential element of interesting stories. In this paper, we operationalize a narratological definition of conflict and extend established narrative planning techniques to incorporate this definition. The conflict partial order causal link planning algorithm (CPOCL) allows narrative conflict to arise in a plan while maintaining causal soundness and character believability. We also define seven dimensions of conflict in terms of this algorithm's knowledge representation. The first three—participants, reason, and duration—are discrete values which answer the "who?" "why?" and "when?" questions, respectively. The last four—balance, directness, stakes, and resolution—are continuous values which describe important narrative properties that can be used to select conflicts based on the author's purpose. We also present the results of two empirical studies which validate our operationalizations of these narrative phenomena. Finally, we demonstrate the different kinds of stories which CPOCL can produce based on constraints on the seven dimensions.**

***Index Terms*—Conflict, narrative, planning.**

I. Introduction

CONFLICT is a key component of interesting stories. Abbott notes that it "is so often the life of the narrative" [1]. Herman *et al.* go so far as to declare it a "minimal condition for narrative" [2], while Brooks and Warren even tell us that "story means conflict" [3]. It serves at least two important functions.

- Conflict structures the discourse: Traditionally, stories introduce a central conflict early. This jars the world out of equilibrium and causes characters to plot toward the climax—the resolution of that conflict. Longer stories are often segmented into chapters based on smaller conflicts.
- Conflict engages the audience: It causes the audience to ask questions and form expectations about the outcome, which propels them forward through the plot [1], [4], even when the outcome is already known [1], [4], [5].

This agreement on the importance of narrative conflict throws into sharp relief the lack of computational literature written on the subject. Narratologists as far back as Aristotle [6] rely on a reader's implicit understanding of conflict. Because machines have no such implicit understanding to fall back on, a formal model of narrative conflict needs to be distilled in order to leverage its important properties in automatically generated stories. Narrative-oriented virtual environments like role-playing games (RPGs), training simulations, and interactive tutoring systems often need to adapt their stories in response to user actions, so these systems stand to benefit from such a model. A computational formalism can also provide important insights for the overall project of studying human narrative cognition.

This paper describes a computational model of plan-based fictional narrative conflict along with a planning algorithm suitable for automatic story generation systems. It also presents seven dimensions that can be used to differentiate one conflict from another—participants, reason, duration, balance, directness, stakes, and resolution. We present the results of two human subject experiments which demonstrate that our formalization of these concepts corresponds to the narratological definitions. Last, we demonstrate how the dimensions of conflict can be used to constrain the algorithm to produce stories with certain properties.

II. Related Work

We survey related work in three important areas: computational narratology, previous narrative generation systems, and the quantitative analysis of stories.

A. Conflict in Computational Narratology

In our previous work, partial order causal link (POCL) plans have been useful data structures for representing stories because they explicitly model action, temporality, and causality [7]. These are key ingredients for a story's fabula, as described by narratologists [4], [8], and they constrain story generators to produce stories which have a logical causal progression from one event to the next. POCL plans have been used to model suspense and surprise [9], task learning in a narrative environment [10], salience of events in a narrative [11], and other phenomena.

Previous work by Riedl and Young [12] further constrained POCL planning to produce more believable stories in which

Manuscript received July 25, 2012; revised February 27, 2013 and May 11, 2013; accepted July 24, 2013. Date of publication October 07, 2013; date of current version September 11, 2014. This work was supported by the U.S. National Science Foundation (NSF) under Award IIS-0915598. The opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the NSF.

S. G. Ware was with the Department of Computer Science, North Carolina State University, Raleigh, NC 27695 USA. He is now with the University of New Orleans, New Orleans, LA 70148 USA (e-mail: sgware@ncsu.edu).

R. M. Young and D. L. Roberts are with the Department of Computer Science, North Carolina State University, Raleigh, NC 27695 USA.

B. Harrison was with the Department of Computer Science, North Carolina State University, Raleigh, NC 27695 USA. He is now with the Georgia Institute of Technology, Atlanta, GA 30332 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCIAIG.2013.2277051

characters are seen to pursue their personal goals while the story as a whole moves toward goals defined by the author. In this formalism, every character's plan must succeed, and Riedl and Young identified the need to further extend the model to allow for conflict via thwarted and failed plans.

Many narratological definitions of conflict focus on the central role of actions, plans, and the thwarting of plans [2], [13], [14]. Egri [13] and Dibell [14] distinguish conflict from the more general notion of tension by specifying that conflict is a property of thwarted intentional actions (i.e., plans). Tension is the general sense of opposition between forces, so while conflict delivers tension, it is not the only source. In terms of the belief desire intention (BDI) framework [15], conflict arises from intentions, not desires. Here our work differs from others like Szilas [16], who defines conflict as opposition between a character's actions and moral principles—something we would label as tension. The definition we have chosen to operationalize is this: conflict occurs when a character forms a plan that is thwarted by another event in the story, or would have been thwarted if the event had succeeded. The thwarting event can be part of another character's plan (external conflict), part of the same character's plan for a different goal (internal conflict), or an accident or force of nature (environmental conflict).

B. *Conflict Generation*

As far back as Talespin [17] 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 preestablished plot. While stories may branch based on user choices, the content of these branches is usually also preestablished at design time.

This method was generalized by systems such as Universe [19] and Mexica [20], which combined prescripted 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 utilize prescripted 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 and Witten [21] generated conflict by casting the protagonist and the antagonist of a story as competing agents in a zero-sum game and using the minimax game playing algorithm to generate stories. This is a principled approach to conflict generation, but it oversimplifies the antagonist's motivation. The antagonist is not simply a malevolent force to make trouble for the protagonist, but a character with its own goals that should thwart the protagonist only when those goals require it to do so.

Barber and Kudenko [22] created 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. However, since GADIN's dilemmas arise and get resolved immediately, it is difficult to model the thematic and extended conflicts that provide important macrostructural features of a story.

Teaching conflict resolution strategies was a focus for the Fearnot! [23] and SIREN [24] narrative systems. Both used a model of conflict based on organizational psychology research, as opposed to our narratology-inspired model, so they tended to emphasize different aspects. For example, conflict resolution games designed to teach real-world skills tend to focus on cooperation and compromise, whereas fictional conflicts are more often resolved by competition or trickery.

Our model leverages threatened goals in story plans to represent conflict. At least two previous systems have used similar models. Carbonell [25] described an early model of plans and counterplans, but it requires detailed situational information such as when and how goals interact, plan scripts, counterplan scripts, and domain-specific heuristics for selecting which strategies to apply. By extending the general framework of classical AI planning we hope to reduce the authoring burden and benefit from the reusability of narrative planning domains.

Before our model, Gratch and Marsella [26], [27] used threatened goals to recognize conflict and produce appropriate emotional responses in affective virtual agents. Our model is highly compatible with theirs for that reason, but can be differentiated because it is explicitly focused on plot generation, is evaluated by human subjects in a narrative context, and provides a formal planning algorithm. Since the development of our model, Battaglino and Damiano [28] have also modeled conflict as threatened goals.

C. *Analyzing Conflict With Metrics*

Yannakakis [29] provided a survey of research that measures human perceptions of story properties like fun and flow in the context of video games. Peinado and Gervás [30] collected four metrics from human readers evaluating the quality of stories produced by their ProtoPropp system: linguistic quality, coherence, interest, and originality. Our approach to the seven dimensions of narrative conflict differs from these because we measure story properties apart from their effects on the reader. Our dimensions of conflict answer who? what? when? and how? questions; they are designed so that readers can agree on their values even when they disagree on how fun or interesting a conflict is.

At least five story systems have reasoned about conflict quantitatively. IDtension [16] assigned each action a "conflict value" for the degree to which a character is forced to act against its moral principles. MEXICA [20] measured the tension a reader perceives at each world state, allowing the system to craft a pattern of rising and falling action. Zambetta *et al.* [31] specified the ideal amount of conflict in a story as a system of differential equations that simulate an arms race scenario. These approaches are helpful as high-level control for the pace of a story, but cannot reason about the individual motivations of the participants. Gratch and Marsella [26], [27] used the quantifiable metrics of character utility and the likelihood of plan success to differentiate between conflicts. Our four continuous dimensions of conflict—balance, directness, stakes, and resolution—are also defined in terms of utility and likelihood of success, which further demonstrates the compatability of our model of plot generation with their model of affect.

The AI Director of the game series *Left 4 Dead* [32] moderated the intensity of conflicts by controlling the number and frequency of enemies, distribution of powerups, and geography of levels. The director monitored metrics like the player's health and accuracy to measure stress level and create a series of peaks and valleys in intensity that are similar to popular narratives in its domain. We provide a model which can generalize to many domains, but *Left 4 Dead* is an excellent example of how quantifiable metrics can be used to guide the construction of a story according to the author's rhetorical purpose.

III. The CPOCL Model of Conflict

The definition of conflict we have chosen is inherently tied to intentionality and planning. Our model and algorithm extend the work of Riedl and Young's intentional POCL (IPOCL [12]), which in turn extended the POCL model and algorithm [33]. Weld [34] provides a detailed explanation of STRIPS-style planning and the POCL algorithm; we reproduce only those definitions which are needed to explain our model.

Throughout this paper, we use a running example set in the Wild West which includes four characters: Hank (the rancher), Timmy (Hank's son), Carl (the shopkeeper), and William (the sheriff). While the names of the characters were not intended to provide mnemonic cues for their roles in the story, they were the names used in the experiments described in Sections VII and VIII. Since they appear in the screenshots of the experimental materials, they are preserved here for continuity.

A. POCL and IPOCL Plans

A classical planner solves this problem: given an initial world state, a goal, and a set of action templates called operators which have preconditions and effects, find a sequence of steps (concrete instances of operators) that can be applied from the initial state to reach the goal. Fig. 1 gives a small example domain and problem in which the elements of a classical planning problem can be seen.

The POCL family of planners produces a partially ordered sequence of steps and a set of causal links that describe how the preconditions of every step are satisfied by earlier steps.

Definition 1: A causal link is denoted $s \xrightarrow{p} u$, where s is a step with some effect p and u is a step with some precondition p. Step u's causal parents are all steps s such that there exists a causal link $s \xrightarrow{p} u$. A step's causal ancestors are its causal parents in the transitive closure of the parent relation.

A causal link explains how a precondition of a step is met. In other words, p is true for u because s made it so. The example plan in Fig. 2 shows a causal link from step 3 (where Hank steals the medicine) to step 8 (where Hank uses the medicine to heal Timmy) that explains how Hank got the medicine.

During planning, it may be possible for a third step to undo p before it is needed.

Definition 2: A causal link $s \xrightarrow{p} u$ is threatened by a step t iff t has the effect $\neg p$, t can occur after s (which establishes p), and t can occur before u (which needs p).

This notion of threatened causal links is an explicit representation of conflict, but it was originally devised for removing conflict from plans in order to guarantee their success. Our model works by strategically maintaining certain threatened causal links while still guaranteeing a plan's causal soundness and success, but in order to describe it, we need a model of intentionality.

Intentional planning further constrains classical planning. It distinguishes between the author's goal, which must be true by

Problem

```
Initial State:
alive(Hank)∧alive(Timmy)∧alive(William)∧
alive(Carl)∧parentof(Hank,Timmy)∧sheriff(William)∧
armed(Hank)∧armed(William)∧belongsto(Meds,Carl)∧
has(Carl,Meds)∧at(Hank,Ranch)∧at(Timmy,Ranch)∧
at(William,Saloon)∧at(Carl,GenStore)∧
∀?p alive(?p)→{intends(?p,alive(?p))∧¬dying(?p)}
∧intends(Hank,alive(Timmy))∧
intends(Hank,¬dying(Timmy))

Goal:
dying(Timmy)∨¬dying(Timmy)
```

Domain

```
snakebite(?p) "?p is bitten by a snake."
P:  alive(?p)
E:  dying(?p)∧intends(?p,¬dying(?p))∧
    ∀?r parentof(?r,?p)→intends(?r,¬dying(?p))
C:  ∅

travel(?p,?f,?t) "?p travels from ?f to ?t."
P:  alive(?p)∧at(?p,?f)
E:  at(?p,?t)∧¬at(?p,?f)
C:  ?p

take(?p,?i,?h,?l) "?p takes item ?i from ?h at ?l."
P:  alive(?p)∧at(?p,?l)∧has(?h,?i)∧at(?h,?l)
E:  has(?p,?i)∧¬has(?h,?i)∧
    belongsto(?i, ?h)→{ intends(?h,has(?h,?i))∧
    ∀?s sheriff(?s)→intends(?s,¬alive(?p)) }
C:  ?p

heal(?h,?p,?l) "?h heals ?p at ?l."
P:  has(?h,Meds)∧alive(?h)∧at(?h,?l)∧
    dying(?p)∧alive(?p)∧at(?p,?l)
E:  ¬dying(?p)∧¬has(?h,Meds)
C:  ?h, ?p

shootout(?s,?p,?l) "?s and ?p have a shootout at ?l."
P:  armed(?s)∧alive(?s)∧at(?s,?l)∧
    alive(?p)∧at(?p,?l)
E:  ¬alive(?p)
C:  ?s
```

Utility and Likelihood Functions

```
utility(?c):
    u = 0;
    if(alive(?c) ∧ ¬dying(?c))
        u += 0.7;
    if(∃?i belongsto(?i,?c))
        if(has(?c,?i))
            u += 0.3;
    else
        u += 0.3;
    if(∃?h parentof(?c,?h))
        u = (0.4 u) + (0.6 utility(?h));
    return u;

π(step):
    if(step.operator = snakebite)
        return 0.05;
    if(step.operator = take)
        if(¬alive(step.?h))
            return 1;
        if(armed(step.?t) ∧ ¬armed(step.?h))
            return 0.8;
        if(¬armed(step.?t) ∧ armed(step.?h))
            return 0.2;
        else
            return 0.5;
    if(step.operator = shootout ∧ armed(step.?p))
        return 0.5;
    return 1;
```

Fig. 1. Example CPOCL problem and domain. Each operator has preconditions P, effects E, and consenting characters C. Example functions for calculating utility and π are given. This domain uses conditions, disjunctions, and quantifiers as described by Weld [34].

the end of the story, and the goals of individual characters which can change during the plan. Intentional planning operators are annotated to describe which characters must consent to a step before it is taken. An intentional plan is only valid if every step either achieves some character goal or is the causal ancestor of such a step. In the example domain, the person taking an item, `?p`, must consent to the `take` step, and it can only be part of a plan if it achieves one of `?p`'s goals or causally contributes to achieving a goal later.

The IPOCL algorithm groups a plan's steps into frames which explain how they relate to character goals.

Definition 3: An intention frame is a five-tuple $\langle c, g, m, \sigma, T\rangle$ where c is a character, g is some goal that c intends to make true, m is the motivating step which has effect $\texttt{intends}\ (c, g)$, σ is the satisfying step which has effect g, and T is a set of steps taken by c to achieve g. T is called the character's subplan to achieve goal g. The satisfying step σ is in T, and all steps in T must occur after m and be causal ancestors σ.

Some steps, called happenings, do not require the consent of any characters. To simplify future definitions, each happening in a plan is placed in its own intention frame that is intended by a special character called Fate.

Intentional planning constrains classical planning to create more believable stories, but these constraints may be too narrow. For IPOCL, if a character forms a subplan for a goal then that subplan must succeed. In order to model conflict, we need to allow for thwarted and partially executed plans.

B. CPOCL Plans

In a conflict POCL (CPOCL) plan, every step has a boolean flag that is true if the step is an executed step and false if it is a nonexecuted step. An executed step is one which will be executed at some point in the story, whereas a nonexecuted step is one which will never be executed. A nonexecuted step which is part of a character's subplan is a step that character intended to take but was not able to. In Fig. 2, Hank intended to heal Timmy but that step never occurs.

The existence of nonexecuted steps implies a new constraint on causal links: a causal link can never have a nonexecuted step as its tail and an executed step as its head. In other words, a step which never occurs cannot satisfy the preconditions of a step which does occur.

Another problem arises when a character adopts a goal g that is already satisfied. If the character wishes the goal to remain true, we use for the satisfying step a dummy persistence step, which has one precondition g and one effect g. Persistence steps are nonexecuted and ordered to occur at the end of the plan. The need for persistence steps can be avoided in intentional planning, but they are essential to conflict planning because they create causal links that represent how a character intends for a goal to remain true.

We can now formally define CPOCL plans and conflict.

Definition 4: A CPOCL plan P is a five-tuple $\langle S, B, O, L, I\rangle$, where S is a set of executed and nonexecuted steps, B is a set of binding constraints on the free variables in S, O is a partial ordering of the steps in S, L is a set of causal links joining the steps in S, and I is a set of intention frames describing the subplans in S.

The extensions of nonexecuted and persistence steps allows a CPOCL plan to retain certain threatened causal links without violating the causal soundness of the plan or preventing it from reaching the author's goal.

Definition 5: A conflict in a plan $P = \langle S, B, O, L, I\rangle$ is a four-tuple $\langle c_1, c_2, s \xrightarrow{p} u, t\rangle$ such that:

- c_1 and c_2 are characters, possibly the same;
- there exists a causal link $s \xrightarrow{p} u \in L$ threatened by t (henceforth, such a link is a conflict link);
- there exists an intention frame $f_1 = \langle c_1, g_1, m_1, \sigma_1, T_1\rangle$ such that $u \in T_1$;
- there exists an intention frame $f_2 = \langle c_2, g_2, m_2, \sigma_2, T_2\rangle$ such that $t \in T_2$ and $f_1 \neq f_2$;
- either t or u (or both) is a nonexecuted step.

In other words, a character c_1 forms a subplan. Some causal link in that subplan is threatened by step t, and if t succeeds, then c_1's subplan will fail. If c_1's subplan succeeds, then t must have failed. It is also possible that both the subplan and t will fail. This definition aligns with narratological descriptions [2], [3], [13], [35]. Internal conflict occurs when $c_1 = c_2$ and a character thwarts its own plans. External conflict with other characters occurs when $c_1 \neq c_2$. Conflict with the environment occurs when T_1 or T_2 contains only a happening and is thus intended by Fate.

In Fig. 2, the causal link from step 3 to step 8 that describes how Hank has the medicine is threatened by step 7 where Carl takes the stolen medicine back from Hank. Since step 8 is nonexecuted, this threat does not prevent the plan from reaching the author's goal. This example demonstrates how our model allows for the representation of failed plans and conflict in the intentional planning framework.

IV. Dimensions of Conflict

Our model is intentionally broad in order to cover the diverse phenomenon of conflict. In order to differentiate one conflict from another, we distilled seven dimensions from narratological sources which describe various important properties of conflict. The first three dimensions—participants, reason, and duration—have discrete values which can be observed directly in a CPOCL plan. The last four—balance, directness, stakes, and resolution—have continuous values which require some additional context information. These four dimensions are measured from some character's point of view; e.g., $\text{balance}(c_1)$ expresses how balanced a conflict involving character c_1 is from c_1's point of view.

In this section, we assume the conflict as defined in Definition 5. That is, character c_1 intends step u as part of subplan T_1, and character c_2 intends step t as part of subplan T_2, and step t threatens some causal link in c_1's subplan. Note that c_2 can be Fate. We also rely on two additional functions with range $[0, 1]$ that are already provided by many kinds of narrative systems.

- $\pi(T)$ measures how likely some sequence of actions T is to succeed. Many systems, especially role playing games, involve statistical models of how likely an action is to succeed based on chance (e.g., a dice roll).
- $\text{Utility}(c, T)$ measures how satisfied character c is with the state of the world after the sequence of actions T occurs. $\text{Utility}(c, \emptyset)$ is the character's utility before the conflict

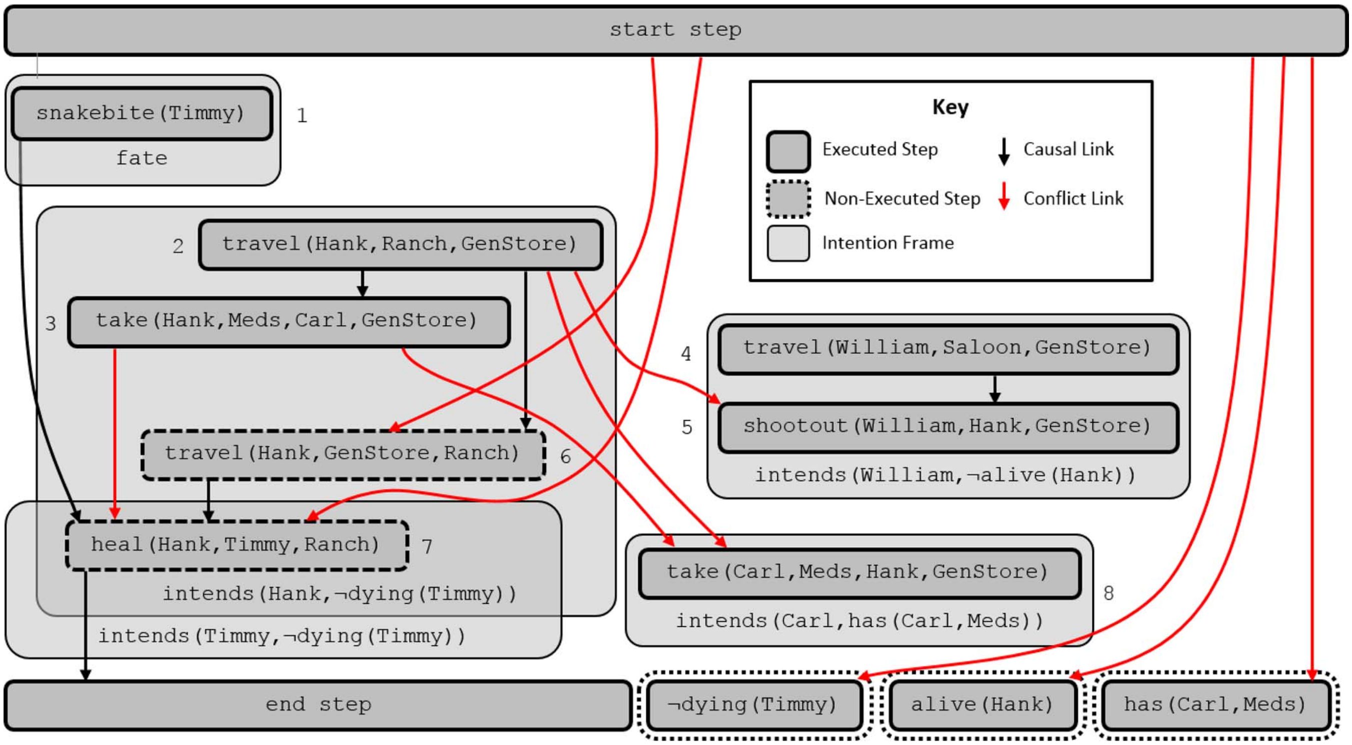


Fig. 2. Example CPOCL plan, which is a solution to the problem in Fig. 1. Numbers next to the steps indicate a valid total ordering. Persistence steps are named for their persisting goals. Twenty eight causal links from the start step and six persistence steps which are not the tail of a conflict link have been omitted to avoid clutter. In this story, Timmy gets bitten by a snake, which prompts Hank to travel to the general store and steal medicine from Carl. Hank plans to return to the ranch and heal Timmy, but this plan is thwarted by sheriff William. In answer to the theft, William travels to the general store and starts a shootout that kills Hank. Finally, Carl takes the medicine back. For a more complete analysis of conflicts involving Hank in this story, see Section IX.

begins. This function might correspond to a player's score or level.

The story in Fig. 2 contains a conflict between Hank and William that culminates in a shootout. A running analysis of that conflict will provide examples throughout this section.

A. Participants

The participants of a conflict are c_1 and c_2, the two characters associated with the conflicting intention frames. In the example conflict, the participants are Hank and William.

B. Reason

The reason of a conflict is the condition which makes the two subplans incompatible—the label of the conflict link. Textually, it can be expressed as "c_1 intends step u, which requires p, but c_2 intends step t, which would cause $\neg p$." For example, Hank intends to heal Timmy, which requires Hank to be alive, but William intends to have a shootout with Hank, which would cause Hank to die.

C. Duration

The duration of a conflict is the span of time during which both participants intend their incompatible subplans. The steps of a CPOCL plan are partially ordered, so to calculate duration, some total ordering O is chosen. Let $\mathrm{index}(s, O)$ 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 has 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. Time_0 is the initial state of the story, occurring after the placeholder start step and before the first step (i.e., the step with index 1). Time_1 is the state after the first step has occurred, time_2 is the state after the first two steps have occurred, etc. We define the duration of a conflict as the number of states during which c_1 intends T_1 and c_2 intends T_2. In order to determine this, we need to know when intention frames begin and end. The beginning is simply the state after the motivating step; however, detecting the end is more complicated.

The end of an intention frame can be thought of as 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 nonexecuted, the frame ends after the last executed step. One important exception to this rule exists: if the first nonexecuted step in a frame is step u of a conflict (the head step of a threatened casual link), then the intention frame ends after step t (the threatening step). The reason for this exception lies in the nature of conflict: if a character abandoned a plan because it was thwarted, it should intend the plan up until the time when the plan gets thwarted.

Let the function $\Omega(f)$ return the index of the earliest state by which intention frame f has ended. Also recall 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

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

An example will help to make this clearer. Using the total ordering given in Fig. 2, we see that Hank forms his plan to heal Timmy right after the snakebite, at time_1. Everything goes well until Hank steals the medicine from Carl at time_3. This prompts William to plot Hank's death, so a conflict begins. It continues until time_5 when William has a deadly shootout with Hank. Thus, the conflict lasts for $5 - 3 = 2$ states.

D. Balance

Balance measures the relative likelihood of each side in the conflict to succeed (regardless of the actual outcome). Before discussing the formula for balance, we need to consider which steps from the story are relevant.

Let T_1' be the steps in T_1 plus their causal ancestors, such that the time index of all steps in T_1' is higher than the start time of the conflict. T_1' is only those future steps which need to occur to carry out the rest of c_1's subplan. Let T_2' be the same to T_2. These sets are also used in the definitions for stakes and resolution.

When the head step u of a conflict link $s \xrightarrow{p} u$ is a persistence step, then $\mathrm{balance}(c_1)$ is simply $1 - \pi(T_2')$ and $\mathrm{balance}(c_2) = \pi(T_2')$. In other words, when c_1 wants some fact to remain true, the balance of the conflict for c_1 is 1 minus the probability that the opponent will succeed. A more general formula is needed when not dealing with persistence steps. Assuming that one side or the other will prevail

$$\mathrm{balance}(c_1) = \frac{\pi(T_1')}{\pi(T_1') + \pi(T_2')}.$$

The range of $\mathrm{balance}$ is $[0, 1]$. If c_1 is likely to prevail, i.e., $\pi(T_1')$ is close to 1, then balance is high for c_1. If the opponent is more likely to prevail, then balance is low for c_1.

In the example conflict, T_1' is the remaining portion of Hank's subplan: to return to the ranch and heal Timmy. Based on the example function provided in Fig. 1, $\pi(T_1') = 1$. William's subplan T_2' is to travel to the general store and shoot Hank. $\pi(T_2') = 0.5$ because the shootout has only a 50% probability of succeeding, since Hank also has a gun. Thus, $\mathrm{balance}(\mathrm{Hank}) = 0.667$, which is skewed in Hank's favor. This makes sense when we consider that Hank's plan has no chance of failing as long as no one interferes, whereas William's plan might fail even if he starts the shootout.

A value of 0.5 corresponds to a most balanced conflict from the point of view of the author, because both participants are equally likely to succeed.

E. Directness

Directness measures how close the participants are to one another

$$\mathrm{directness}(c_1) = \frac{\sum_{i=1}^{n} \mathrm{closeness}_i(c_1, c_2)}{n}.$$

For simplicity, only two types of closeness are measured in this domain: physical closeness and family closeness. Other types can be measured, such as interpersonal closeness [36], which is high when characters carry out their own plans and low when characters accomplish their plans vicariously through other characters. This formula can also be made a weighted average based on genre expectations. The range of directness and each form of closeness is $[0, 1]$.

Hank and William are in the same physical space, but they are not family. Hence, $\mathrm{directness}(\mathrm{Hank}) = 0.5$. Story C in Section IX demonstrates a conflict between Hank and his son Timmy. Because the characters are in the same space and related to one another, the conflict has the maximum closeness of 1 in this domain.

F. Stakes

Stakes is the difference between how high a participant's utility will be if she prevails and how low it will be if she fails (which can be estimated by how low it will be if her opponent prevails)

$$\mathrm{stakes}(c_1) = |\mathrm{utility}(c_1, T_1') - \mathrm{utility}(c_1, T_2')|.$$

The range of stakes is $[0, 1]$. Two factors influence this formula: how much can be gained and how much can be lost. Situations which are high risk (failure results in a low utility) or high reward (success results in a high utility) have medium stakes, while situations which are both high risk and high reward have high stakes. Like balance, stakes is measured regardless of the actual outcome.

Hank's conflict with William has the maximum stakes; $\mathrm{intensity}(\mathrm{Hank}) = 1$. If Hank succeeds, he and his son will both live, which would yield $\mathrm{utility}(\mathrm{Hank}, T_1') = 1$. If he fails, he will die and his son will be left to die of the snakebite, which would yield $\mathrm{utility}(\mathrm{Hank}, T_2') = 0$.

G. Resolution

Resolution measures the change in utility a participant experiences after a conflict ends. Let E be the set of executed steps from T_1' and T_2'. In other words, E is how the conflict actually plays out. It may contain some steps from both subplans, but it cannot contain all steps from both subplans (because they conflict)

$$\mathrm{resolution}(c_1) = \mathrm{utility}(c_1, E) - \mathrm{utility}(c_1, \emptyset).$$

The range of resolution is $[-1, 1]$.

Timmy was already dying when the conflict began, so $\mathrm{utility}(\mathrm{Hank}, \emptyset) = 0.4$. Hank dies in the shootout, so $\mathrm{utility}(\mathrm{Hank}, E) = 0$. Thus, $\mathrm{resolution}(\mathrm{Hank}) = 0 - 0.4 = -0.4$, indicating that the conflict ended poorly for Hank.

V. The CPOCL Algorithm

The CPOCL planning algorithm is given in Algorithm 1. It produces CPOCL plans as described in the previous sections. This algorithm extends the classical POCL algorithm [33] and incorporates intentional planning similarly to the IPOCL algorithm described by Riedl and Young [12]. We assumed a function $\mathrm{MGU}(p, q)$ which returns a set of variable bindings to make $p = q$. Line numbers in the descriptions below correspond to lines in Algorithm 1.

The POCL family of algorithms are a kind of refinement search [37]. A partial plan is annotated with flaws to indicate how it is incomplete. These flaws are iteratively fixed until a

Algorithm 1 The CPOCL (Conflict Partial Order Causal Link) Planning Algorithm

CPOCL ($\Pi = \langle S, B, O, L, I \rangle$, Λ, F)

Π is a plan, initially the null plan, with steps S, variable bindings B, ordering constraints O, causal links L, and intention frames I; Λ a set of operators; F a set of flaws, initially open precondition flaws for unsatisfied preconditions of the end step and unsatisfied intention frame flaws for start step effects like $intends(c, g)$.

1: **Termination:** If B or O is inconsistent, fail. If $F = \emptyset$ and Π has no orphans, return Π. If orphans exist, fail.
2: **Plan Refinement:** Choose a flaw $f \in F$. Let $F' = F - \{f\}$.
3: **Goal Planning:** If f is open precondition flaw $f = \langle s_{need}, p \rangle$, let s_{add} be a step $\langle P, E, C \rangle$ such that $p \in E$.
4: Choose s_{add} in one of two ways:
5: **Reuse:** Choose s_{add} from S.
6: **New Step:** Create s_{add} from an operator in Λ with effect p. Let $S' = S + \{s_{add}\}$.
7: For each precondition pre of s_{add}, add new open precondition flaw $\langle s_{add}, pre \rangle$ to F'.
8: Mark s_{add} as non-executed.
9: **Link:** Create causal link $l = s_{add} \xrightarrow{p} s_{need}$. Let $L' = L + \{l\}$, $B' = B \cup \mathrm{MGU}(e, p)$, $O' = O + \{s_{add} < s_{need}\}$.
10: **Execution Marking:** If s_{need} is executed, mark s_{add} and all its causal ancestors as executed.
11: **Happening Frame:** If $P = \emptyset$, create new intention frame $r = \langle fate, \emptyset, s_{add}, s_{add}, \{s_{add}\} \rangle$. Let $I' = I + \{r\}$.
12: **New Frames:** For each effect of s_{add} like $intends(c, g)$:
13: Create new intention frame $r = \langle c, g, s_{add}, \emptyset, \emptyset \rangle$. Let $I' = I + \{r\}$.
14: Add new unsatisfied intention frame flaw $\langle r \rangle$ to F'.
15: **Intent Flaws:** For each intention frame $r = \langle c, g, \sigma, m, T \rangle \in I'$:
16: If $s_{add} \notin T$ and $s_{need} \in T$ and $c \in C$ for s_{add}, add new intent flaw $\langle s_{add}, r \rangle$ to F'.
17: **Threat Resolution:** If f is threatened causal link flaw $f = \left\langle s \xrightarrow{p} u, t \right\rangle$, choose how to prevent the threat:
18: **Promotion:** Let $O' = O' + \{t < s\}$.
19: **Demotion:** Let $O' = O' + \{u < t\}$.
20: **Restriction:** Add bindings to B' which cause the threatening effect of t not to unify with p.
21: **Satisfaction:** If f is unsatisfied intention frame flaw $f = \langle r = \langle c, g, m, \emptyset, T \rangle \rangle$, let s_{sat} be a step with effect g.
22: Choose s_{sat} the way s_{add} is chosen (**Reuse** or **New Step**) or by **Persistence**.
23: **Persistence:** Make a persistence step $s_{sat} = \langle \{g\}, \{g\}, \{c\}, false \rangle$. Let $O' = O + \{s_{sat} = s_{end}\}$.
24: Let $T' = T + \{s_{sat}\}$. Let $r' = \langle c, g, m, s_{sat}, T' \rangle$. Let $I' = I - \{r\} + \{r'\}$.
25: **Intent Planning:** If f is an intent flaw $f = \langle s_{orphan}, r = \langle c, g, m, \sigma, T \rangle \rangle$, choose how to handle s_{orphan}:
26: **Inclusion:** Let $T' = T + \{s_{orphan}\}$. Let $r' = \langle c, g, m, \sigma, T' \rangle$, $I' = I - \{r\} + \{r'\}$, $O' = O + \{m < s_{orphan}\}$.
27: For each causal link $s \xrightarrow{p} s_{orphan} \in L$, if $c \in C$ for s, add new intent flaw $\langle s, r' \rangle$ to F'.
28: **Exclusion:** Do nothing.
29: **Threat Detection:** If any casual link $l \in L'$ is threatened by step $\theta \in S'$ and l is not a conflict link,
30: Add new threatened causal link flaw $\langle l, \theta \rangle$ to F'.
31: **Recursive Invocation:** Call CPOCL ($\Pi' = \langle S', B', O', L', I' \rangle$, F', Λ).

flawless (and thus, complete) plan is found or the algorithm fails. The root of the search space is the null plan.

Definition 6: A start step s has no preconditions, and effects equivalent to the initial state of the planning problem. An end step e has no effects, and a precondition for each literal which must be true in the goal state. Start and end steps must be executed steps. A planning problem's null plan is the partial plan $P = \langle \{s, e\}, \emptyset, \emptyset, \emptyset, \emptyset \rangle$.

A refinement planner defines a set of flaws and ways to repair them.

Definition 7: An open precondition flaw indicates that some precondition of a step has not yet been met by a causal link. It is a 2-tuple $\langle s_{\mathrm{need}}, p \rangle$, where s_{need} is some step in S and p is a precondition of s such that no causal link in L has s_{need} as its head and p as its label.

Open precondition flaws are repaired by adding a new causal link to L which has s_{need} as the head (lines 3–10). The tail of the new link can be either a step already in S (line 5) or a new step created from an operator and added to S (lines 6–8). Adding new steps to S may require adding new open precondition flaws (line 7).

When a new step is added to the plan, it is initially marked as nonexecuted (line 8). If a causal link is created from a nonexecuted step to an executed step, the tail step and all its causal ancestors must then be marked as executed (line 10).[1] This ensures that nonexecuted steps are never used to satisfy the preconditions of executed steps.

When a happening is added to the plan, it is placed in its own intention frame whose actor is Fate (line 11).

When a step with an effect like `intends` (c, g) is added to the plan, a new intention frame is created with that step as the motivating step (lines 12–14). CPOCL must later choose a satisfying step to explain how character goal g gets fulfilled. This need translates into a flaw.

[1]A complete plan will have only those steps marked as executed that must occur to achieve the goal. This follows the least commitment paradigm of POCL planners. It may be possible to mark additional steps as executed without making the plan unsound, and any system using CPOCL is free to do so.

Definition 8: An unsatisfied intention frame flaw indicates that a satisfying step has not yet been chosen for an intention frame. It is a 1-tuple $\langle f \rangle$, where f is some intention frame.

After a satisfying step is chosen (lines 21–24), the frame must be populated with all the steps that the character takes in pursuit of the goal.

When a new causal link is created, it may link a step outside an intention frame ($\notin T$) to a step inside an intention frame ($\in T$). This might indicate that the outside step was taken in pursuit of the frame's goal. If so, the outside step needs to be included in the frame. This need is represented as a flaw.

Definition 9: An intent flaw occurs when a causal link $s \xrightarrow{p} u$ exists such that, for some intention frame $r = \langle c, g, m, \sigma, T \rangle$, $s \notin T$, $u \in T$, and c is a character who must consent to s. It is a 2-tuple $\langle s, f \rangle$, where s is the step which may need to be included in frame f.

Intent flaws can be solved by adding the step to the frame (lines 26–27) or by ignoring the flaw (line 28). It is necessary to consider ignoring the flaw to ensure that valid plans are not missed in the search process, however this decision creates an orphan. Riedl and Young give a full account of dealing with orphans [12].

Definition 10: A threatened causal link flaw indicates that the condition established by a casual link may be undone before it is needed. It is a 2-tuple $\langle s \xrightarrow{p} u, t \rangle$, where $s \xrightarrow{p} u$ is a causal link in L, and t is a step in S which threatens it.

Threatened causal links are fixed by preventing the ordering $s < t < u$ (lines 18–19) or by adding bindings to B which prevents the threatening effect of t from logically unifying with $\neg p$ (line 20). Note that threatened causal link flaws are not added for conflict links because they do not need to be repaired (line 29). In fact, since conflict is a desirable property of stories, they probably should not be repaired.

VI. Characterizing the CPOCL Solution Space

In order to distinguish CPOCL, we must consider a classical POCL planner and an intentional (I)POCL planner that does not support conflict. All define a search through the space of partial plans, but their solutions contain different data structures (i.e., intention frames and nonexecuted steps). In order to compare the solutions across planners, we consider a plan as only a partially ordered set of executed steps.

POCL planners find all plans which are guaranteed to reach the goal from the initial state. Intentional planning restricts this search space to plans in which every nonhappening is taken in service of some character goal. In an IPOCL solution, either a character never attempts to achieve a goal or it succeeds in achieving it.

CPOCL imposes an additional restriction on intentional planning: every motivating step must have an associated intention frame. In other words, if a character adopts a goal, it must attempt to achieve it. However, unlike in IPOCL, a character can fail to achieve that goal. The CPOCL search space is narrower than POCL's but broader than IPOCL's. In short, a CPOCL plan is guaranteed to achieve the planning problem's goal, contains only steps which are taken in service of some character goal (plus happenings), and contains a subplan to achieve every character goal, some of which may fail. Plans like the one in Fig. 2 cannot be produced by IPOCL, even when we consider only the executed steps. Our extensions to planning are valuable not only because they explicitly represent conflict but also because they expand the space of stories which can be represented.

Note that CPOCL's solution space may include plans with no conflict. If the goal can be achieved without creating any conflict—that is, if a solution can be found with intentional planning alone—these plans can appear in CPOCL's solution space. To avoid this, the termination condition can specify that a plan must contain at least one conflict link. However, this may produce stories in which characters go out of their way to create needless conflict (from the audience's perspective). When a problem can be solved without conflict, it may indicate that the initial state and goal of the story need to be revised rather than the planner.

VII. Validating the Planning Structures

Having described the CPOCL model (which encompasses both the planning structures and the seven dimensions), we wish to demonstrate that it correctly operationalizes our chosen narratological definitions. Sections VII and VIII present the results of two experiments which compare human story analysis to that provided by our model.

A. Notes on Experimental Design

It is essential to point out that these experiments were not designed to prove that our definitions of conflict and the seven dimensions are the only "correct" ones; the correctness of our definitions lies with the authority of the narratological sources from which they were derived [2], [13], [14], [35], [38]. We are only attempting to demonstrate that, given our definitions, we have created a computational model that represents them accurately at the fabula level. For this reason, subjects were given short natural language descriptions of conflict and the dimensions. This was purposely done to minimize any disagreement that might arise because subjects assumed different narratological definitions of conflict.

We created the stories for both experiments, rather than using naturally occurring stories, for several reasons. Creating stories allowed us to control for content in the fabula (number of characters, length of story, etc.). This was especially important in the second experiment, which had to reuse characters, places, items, and actions as much as possible to ensure that subjects did not prefer one story over another because, e.g., they prefer dragons over aliens. Creating stories also allowed us to control the discourse, which is especially important given that our model is currently only one of fabula. We intentionally employed a minimal translation from fabula to discourse which avoided discourse-level techniques such as telling the story out of chronological order, deceiving the reader, creating suspense, etc. Creating stories allowed us to control for content in the text as well. Stories were translated from plans into natural language using simple text templates to minimize the effect of stylistic writing and word choice. We were unable to find a set of naturally occurring stories which met all these control requirements, however when creating these stories we did attempt to use people, places, things, and plot devices consistent with the genre of each story. Leaving aside the need for control, avoiding naturally occurring stories was also important to ensure that subjects were not already familiar with the stories before the experiments.

We intend to extend this model in many ways in our future work, including reasoning about conflict at the discourse and text levels. Once the model has been so extended, we intend to demonstrate its applicability as an analysis tool for naturally occurring stories, but in its current form we seek only to establish that it can represent conflict at the most basic level of the fabula and to establish a clear connection between the independent and dependent variables via strictly controlled experiments.

B. Design of the First Experiment

The first experiment evaluates the plan-based structure of CPOCL, along with the three dimensions that can be directly observed in a CPOCL plan: participants, reason, and duration [39]. Human subjects were given three short stories and asked to list all the conflicts they observed. They also answered "who?" "why?" and "when?" questions for each conflict. The definition of conflict they were given was: "A conflict occurs when a participant plans to do something which might get thwarted later." We felt this was a succinct, easy to understand summary of the action-based definitions of conflict given by Egri [13], Dibell [14], Herman [35], Herman *et al.* [2], and others.

We used the data collected to evaluate two hypotheses:

1) subjects will report conflicts similarly to one another;
2) subjects will report conflicts that are similar to those defined by CPOCL when the stories are modeled as CPOCL plans.

The experiment was conducted online, and subjects were recruited via e-mail and social networking websites. No compensation or incentives were offered. Subjects completed a tutorial to familiarize themselves with the interface, and then each subject reported conflicts for all three stories, which were presented in a random order. The stories took place in three different domains: the American west, a medieval fantasy kingdom, and futuristic outer space. Twenty seven people responded to the survey by finishing one or more stories. Of those, 23 subjects finished all three stories. There were 16 male and 11 female subjects. The most common age range was 26–35. In total, 486 conflicts were reported across the three stories. If a subject reported no conflicts for a story, that subject's data were not included in our analysis for that story.

The interface (shown in Fig. 3) allowed subjects to move backward and forward through time at will. At each moment, they were shown the story up to that point along with thought bubbles for each character (including Fate) that describe the character's current plan. Because CPOCL is a model of story fabula, subjects were intentionally given this god's-eye-view of the story. This avoided the need for clever discourse techniques to inform readers of what each character was planning (such as flashforwards, villain monologs, etc.).

Subjects were asked to list all the conflicts they noticed. A conflict was reported via a point-and-click interface as a 6-tuple $\langle c_1, c_2, s_1, s_2, b, e\rangle$, which was composed of:

- c_1, the first character;
- c_2, the second character;
- s_1, an action from c_1's thought bubble;

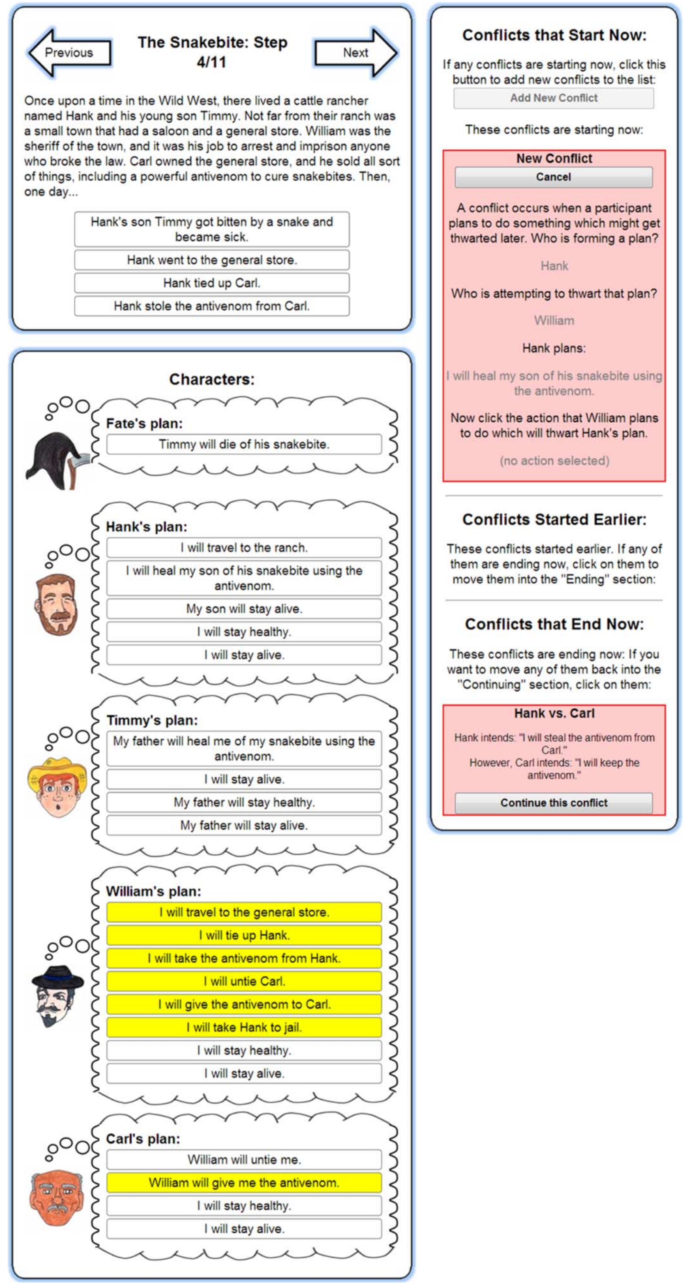


Fig. 3. Interface used to report conflicts in the first experiment. The story domain is a more robust version of Fig. 1. The top box shows the story up to the current moment. The bottom box shows the current plans of all characters (with changes from the previous moment highlighted). The right box allows users to report conflicts. The user above is reporting a conflict between Hank and William. Subjects moved conflicts between the start now, started earlier, and end now groups to report duration.

- s_2, an action from c_2's thought bubble that thwarts c_1's plan;
- b, the time when the conflict begins;
- e, the time when the conflict ends.

This information describes the participants, reason, and duration of each conflict. The order of participants was ignored; in other words $\langle c_1, c_2, s_1, s_2, b, e\rangle = \langle c_2, c_1, s_2, s_1, b, e\rangle$.

C. Intersubject Agreement

Before evaluating CPOCL, we must establish that subjects agree among themselves. For this, we used Fleiss' κ coefficient.

TABLE I
FLEISS' κ (SUBJECT AGREEMENT) FOR THREE STORIES

		exact		overlap	
Story	**subjects**	**# of ?s**	κ	**# of ?s**	κ
Western	25	66	0.31	31	0.49
Fantasy	24	35	0.33	18	0.59
Space	25	49	0.17	21	0.44
Average			**0.27**		**0.51**

Fleiss' κ has the property that $1 \geq \kappa > 0$ if subjects agree (with values closer to 1 representing better agreement), and $\kappa \leq 0$ if subjects disagree. Fleiss' κ assumes three or more subjects are answering multiple-choice questions, so we must interpret our data in this fashion. We can define every possible conflict that could have been reported using the interface as a question that was implicitly answered as "yes" if the subject reported it, or "no" if the subject did not report it. However, this will inflate the accuracy of our results with a preponderance of true negatives and skew the κ calculations. Therefore, we define "all possible conflicts" to mean any conflict that was reported by at least one subject, which was less than 1% of all conflicts that could have been reported using the interface. This will reduce the accuracy of our results but provides a fairer evaluation of the model.

The first column of Table I, labeled "exact," shows the κ values achieved for each story. In all cases, $p < 0.001$. "Subjects" is the number of subjects who finished that story. "# of $?s$" is the number of possible conflicts (i.e., number of questions) to which subjects implicitly answered "yes" or "no."

Many of the conflicts reported by subjects had the same participants, same reason, and overlapping (but not exactly the same) duration. 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 I, labeled "overlap." In all cases, $p < 0.001$. 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 and the κ interpretations given by Landis and Koch [40], we conclude that users demonstrated moderate agreement about which conflicts exist in the three stories, especially when allowing for overlapping durations. However, given the subjective nature of Landis and Koch's interpretations and the vulnerabilities of κ, Section VII-D presents additional and perhaps more convincing evidence of agreement.

D. Subject Agreement With CPOCL

In order to evaluate CPOCL's performance relative to human subjects, we need to establish which conflicts are considered correct out of all the ones reported. For each story, we must choose some threshold θ such that if θ or more subjects reported a conflict, that conflict is defined as correct for that story. Choosing a θ value allows us to evaluate the accuracy of an individual subject. Keeping in mind that our data can be interpreted as a number of questions implicitly answered as "yes" if the subject sees a conflict and "no" if the subject does not see it, we define accuracy to be the number of questions answered correctly divided by the total number of questions answered (conflicts with the same participants, the same topic, and overlapping durations are considered the same).

TABLE II
THRESHOLD VALUES FOR EACH STORY

Story	min θ	max θ	Average Accuracy
Western	12 (48%)	19 (76%)	80%
Fantasy	8 (33%)	21 (88%)	81%
Space	4 (16%)	16 (64%)	80%

TABLE III
CONFUSION MATRICES FOR CPOCL'S PERFORMANCE ON TASK 1

Western Story		Fantasy Story		Space Story	
TP: 8	FP: 6	TP: 4	FP: 4	TP: 6	FP: 3
FN: 0	TN: 17	FN: 0	TN: 11	FN: 0	TN: 12

TABLE IV
CPOCL'S ACCURACY (ACC.), PRECISION (PRE.) AND RECALL (REC.) FOR BOTH TASKS

	Task 1			Task 2		
Story	Acc.	Pre.	Rec.	Acc.	Pre.	Rec.
Western	0.81	0.57	1.00	0.98	0.90	1.00
Fantasy	0.79	0.50	1.00	0.90	0.61	1.00
Space	0.86	0.67	1.00	0.87	0.82	0.86
Average	**0.82**	**0.58**	**1.00**	**0.92**	**0.77**	**0.95**

For each story, we chose the lowest value of θ that maximized the average accuracy of subjects. These θ values are given in Table II in the "$\min \theta$" column. Consider the Western story as an example. If we define a correct conflict as one reported by 12 or more subjects (that is, $\theta = 12$ or 48% of subjects), then the average subject's accuracy in reporting conflicts is 80%. $\theta = 12$ is the lowest value of θ that achieves the highest possible average accuracy of 80%. We could have chosen θ as high as 19 and observed the same average accuracy (given in Table II as "$\max \theta$"), but since many subjects reported exhaustion during the experiment, we chose the lowest θ in order to utilize as much data from subjects as possible.

As Table II shows, there exists a θ value for each story such that the average subject achieves 80% or 81% accuracy. This is further evidence that subjects agree about which conflicts exist.

To our knowledge, there are no other formal models of narrative conflict which can be directly applied to this sort of data. Therefore, we compare CPOCL against both a naive baseline and the performance of the average human subject.

Task 1: For Task 1 we treated CPOCL as a subject and compared the set of conflicts that it defines to those reported by humans. The resulting confusion matrices are shown in Table III. A true positive is a conflict defined by CPOCL that θ or more subjects reported. A false positive is a conflict defined by CPOCL that fewer than θ subjects reported. A false negative is a conflict reported by θ or more subjects that CPOCL does not define. A true negative is a conflict which was neither defined by CPOCL nor reported by θ or more subjects. Summary statistics for Task 1 are presented in Table IV.

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

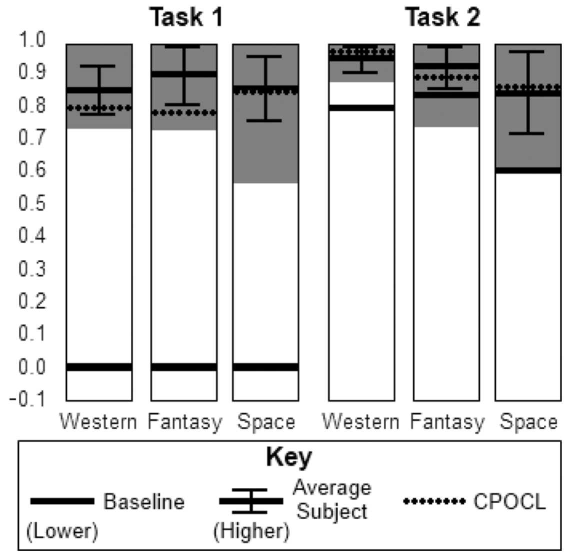


Fig. 4. CPOCL's accuracy versus a naive baseline and the average human subject. The gray region represents the range of accuracy values for individual subjects. The higher solid black bar represents the average subject accuracy ($\pm$ one standard deviation). The lower solid black bar represents the accuracy of the baseline. The dotted line represents CPOCL's accuracy.

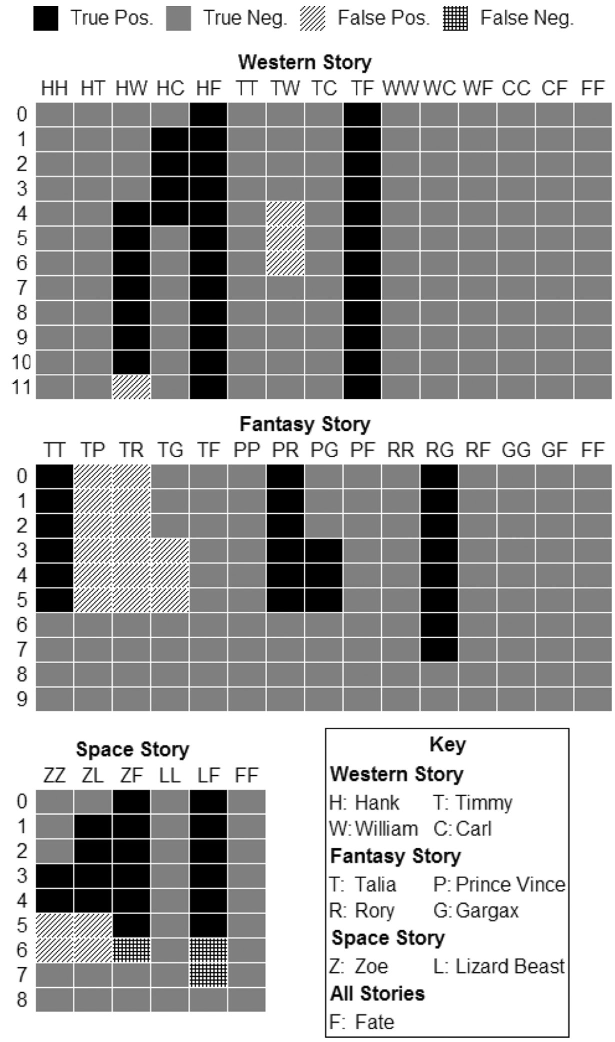


Fig. 5. Visualization of CPOCL's performance on Task 2. For each story, the vertical axis is labeled with the time index of the state. The horizontal axis is every pair of characters.

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 θ or more subjects 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 the average human subject. These results are visualized in Fig. 4. CPOCL does significantly better than random guessing (a very naive baseline). For the Western and Space stories, CPOCL's accuracy is within one standard deviation of the average subject.

Task 2: We can also evaluate the model by testing how well it predicts when a given pair of characters is in conflict. For this task we ask the following question both of human subjects and of CPOCL: for every state, and for every pair of characters, are those characters in conflict in that state?

Fig. 5 visualizes the results. True positives indicate that subjects and CPOCL both answered "yes"; true negatives indicate that both answered "no"; false positives indicate that CPOCL answers "yes," but subjects answered "no"; and false negatives indicate that subjects answered "yes," but CPOCL answered "no." Summary statistics for Task 2 are presented in Table IV.

A naive baseline for this task is to always answer "yes" or "no" to every question. Of those two, answering "no" yields the highest accuracy, and answering "yes" yields the highest precision. We compared CPOCL's performance to these two baseline models, and the results are presented in Table V along with CPOCL's percent improvement over the baseline.

CPOCL did better on Task 2. It always outperformed the baseline, was always within one standard deviation of the average user, and even outperformed the average subject for the Western and Space stories.

E. Discussion

Some conflicts defined by the model were understandably counterintuitive to subjects due to a mismatch between how people think about actions and the knowledge representation of

TABLE V
PERFORMANCE ON TASK 2 RELATIVE TO NAIVE BASELINES: ALWAYS "NO" AND ALWAYS "YES"

	Accuracy			Precision		
Story	**No**	**CPOCL**	**%impr**	**Yes**	**CPOCL**	**%impr**
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%**

a STRIPS-style story domain. An example from the Western story can illustrate this: William intends to take the antivenom from Hank, but Hank intends to travel back to his ranch. This conflict arises because the `take` action requires that both characters be at the same location (in this case, the general store). If Hank travels to his ranch, he will no longer be at the general store and William's `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 readers do not think about steps in terms of their exact mechanics; rather, they think at a more abstract level where going to the general store is the same action no matter where the character is coming from.

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. In the Space story, two characters intend to stay safe from natural disasters, but a volcano erupts which is unsafe for everyone in that area. According to CPOCL, this is the end of a conflict; Fate won because both characters failed to stay safe (which caused them to form new plans to go to safe locations). Subjects recognized these conflicts, but they reported the end as the time when each character had reached a new safe location. This seems like the most natural interpretation of the story, so one future improvement to CPOCL may be the inclusion of durative actions.

Interestingly, no false negatives were reported for any story in Task 1. Even for Task 2, the only false negatives were due to the above disagreement about how long the conflict with the volcano should last, not about who was involved or why. In other words, the conflicts defined by CPOCL are a strict superset of the correct conflicts reported by subjects. This seems reasonable given that CPOCL is a model of fabula and does not yet account for the presentation of conflict at the discourse level. One important direction of future work will be to discover why readers notice some conflicts but not others. Certain threatened causal links are very obvious to subjects, while others (that are not formally or structurally different) seem not to be obvious at all. This may have been due to subject exhaustion—the annotation process was very taxing—or it might be explained by discourse phenomena such as how subjects direct their attention while reading. Some work investigating this is already underway [11].

VIII. Validation of the Dimensions

We designed a second experiment to validate the four continuous dimensions of conflict—balance, directness, stakes, and resolution [36]. Predicting the exact value a subject will report is difficult considering how sensitive these concepts are to subtleties of interpretation. Simply predicting high or low is easier, but success would provide less support for our model. We attempted to reach a middle ground by demonstrating that our formulas can rank four stories in the same order as human subjects. If subjects agree on an ordering, and if that ordering agrees with our predictions, we assume that our formulas can approximate these four dimensions of conflict.

The study was conducted via a web interface in which subjects could drag and drop stories from an initial random order into a sorted order of their choosing. Each subject ranked the same four stories for all four dimensions. Dimensions were presented to each subject in a random order. Subjects were recruited via e-mail, social networking websites, and online message boards. No compensation was offered. Thirty people responded: 19 males and 11 females with the most common age range being 26–35.

An example story is given in Fig. 6. Each story has the same beginning, takes place in the same domain, involves the same characters, and centers around a conflict between the reader and an evil sorcerer. The stories were designed to exhibit a wide range of values for each dimension. For example, if you fight the sorcerer yourself, the conflict is more direct than if you ask the knight to fight for you. When the sorcerer threatens to kill the prince, the conflict has higher stakes than if he makes no threats. This experiment does not require a commitment to specific formulas for $\pi(T)$ and $\mathrm{utility}(a, T)$ as long as those formulas produce the predicted orderings. For example, we assume that the knight is more likely to succeed when he has a sword and armor than when he has just a sword and no armor. It is not necessary to measure the exact difference in π between the two stories.

Introduction
This story takes place in a magical kingdom ruled by a wealthy king. The king has a young son, the prince. You are just a poor farmer, but you are friends with the prince. One day, an evil sorcerer kidnaps the prince! The king offers you a reward if you can get the prince home safely.

Story A
You travel to the city.
You ask a knight to kill the sorcerer.
The knight buys a sharp sword at the market.
The knight travels to the tower.
The knight challenges the sorcerer to a fight to the death.
The sorcerer reveals that he is your father.
The knight defeats the sorcerer.
The prince travels to the city.
The king gives you a bag of gold.
The king makes you a knight.

Fig. 6. Sample story for which subjects reported balance, directness, stakes, and resolution. Each story had the same beginning, but a different middle and end.

The content of the stories was structured so that, given the orderings for each dimension predicted by our formulas, no two stories would appear at the same index for the same dimension. That is, the second most direct story was never ranked second for any other dimension. Subjects were not told of this constraint. It was imposed in an attempt to avoid conflating dimensions. For example, if our formulas assigned the same ordering to balance and stakes and subjects ranked stories in that order for both dimensions, it would be impossible for us to know whether they perceived balance and stakes as two different phenomena or if they were conflating the two.

In order to avoid confusion from vocabulary, the dimensions were not given names in the study. Subjects were simply given short natural language descriptions of each dimension (given in Fig. 7) and asked to sort the stories.

We evaluated two hypotheses:

1) subjects will rank stories similarly to one another;
2) subjects will rank stories similarly to our metrics.

A. Analysis

The data collected from each subject was an ordering of four stories for each dimension. The task of choosing an ordering is similar to classification, but two orderings can still be substantially similar even if they are not exactly identical. This precludes the straightforward application of standard agreement

BALANCE

Rate the stories based on how likely you and your allies are to win out over the sorcerer. If you expect your team to win, rate the story high. If you expect your team to lose, rate it low. Do not consider whether or not you *actually* win. Only rate the stories based on what you expected to happen *before someone gets defeated.*

DIRECTNESS

Rate the stories based on how close you are to the sorcerer. There are many kinds of closeness: physical closeness, emotional closeness, familial closeness, etc. Only consider the distance between *you* and the sorcerer.

STAKES

Rate the stories based on how much is at stake for you. Imagine how bad it will be if the sorcerer wins and how good it will be if you and your allies win. Stories which could end very badly or very well for you should be ranked high. Stories where your happiness is not likely to change very much should be ranked low. Do not consider the *actual* outcome of the story. Only rate the stories based on how much you think is at stake *before someone gets defeated.*

RESOLUTION

Rate the stories based on how much better off you are at the end. How much happier are you at the end of the story than at the beginning? Only consider how *you* have been affected. Do not consider how things *might have been*, only how they actually happened.

Fig. 7. Definitions of the four dimensions given to subjects for the second experiment. Subjects were not given the names of the dimensions, only the descriptions.

measures like Fleiss' κ. To account for similarity between responses, we used Kendall's τ distance [41] to compare orderings. τ counts the number of pairwise differences between two lists. In the study of sorting algorithms, τ distance is sometimes called "inversion count" and is a standard measure of sortedness.

Formally, let $\mathrm{index}(x, S) = 1$ just when x is the first element in ordered set S, $\mathrm{index}(x, S) = 2$ just when x is the second element in ordered set S, etc. Given two ordered sets M and N, an inversion is an ordered pair of elements (x, y) such that $\mathrm{index}(x, M) < \mathrm{index}(y, M)$ and $\mathrm{index}(x, N) > \mathrm{index}(y, N)$. This means that x is ordered before y in M, but x is ordered after y in N. The τ distance between two ordered sets can be expressed as $\tau(M, N)$ and is equal to the number of inversions that exist between M and N. Kendall's τ distance is symmetric, meaning $\tau(M, N) = \tau(N, M)$. When comparing two orderings of length 4, the minimum τ distance is 0 (which means both orderings are the same), and the maximum τ distance is 6 (which means that one is the reverse of the other). If we fix M and choose N at random, assuming that all 24 permutations of the four stories are equally likely, then on average there will be a τ distance of 3 between the two orderings.

To determine the most popular ordering for each dimension based on the data submitted by subjects, we measured the average τ distance for each of the 24 possible permutations of the four stories. For a given dimension of conflict, let $\{p_1, p_2, \ldots, p_n\}$ be the orderings chosen by the n subjects for

TABLE VI
BHATTACHARYYA DISTANCE BETWEEN OBSERVED DISTRIBUTIONS COMPARED TO PERFECT AGREEMENT (PERFECT), RELATIVE AGREEMENT (REL. AGREE), AND DISAGREEMENT (DISAGREE)

Dimension	Perfect	Rel. Agree	Disagree
Balance	0.314	0.108	0.240
Directness	0.255	0.037	0.619
Stakes	0.4658	0.168	0.175
Resolution	0.314	0.040	0.650

that dimension (here, $n = 30$). Let M be all 24 possible orderings of the four stories. For each possible ordering, $m \in M$, its average τ distance is

$$\forall m \in M : \tau_{\mathrm{avg}}(m) = \frac{\sum_{i=1}^{n} \tau(m, p_i)}{n}.$$

Consider four stories, A, B, C, and D, and let $m = \{\mathrm{A\ B\ C\ D}\}$, the first of the 24 permutations in M. To calculate τ_{avg} for m for the dimension of balance, we calculate $\tau(\{\mathrm{A\ B\ C\ D}\}, p_i)$ for all 30 orderings p_i that were reported by the 30 subjects for balance; then we average those 30 values. An ordering's τ_{avg} can be thought of as its average distance from all the reported orderings. When an ordering's τ_{avg} is low, that ordering is more popular—it agrees more with the orderings reported by subjects. If all 30 subjects had reported the same ordering, its τ_{avg} would be 0.

The most popular orderings according to subjects for each dimension are

Dimension	Ordering	τ_{avg}
Balance	C D A B	1.26667
Directness	B A C D	0.56667
Stakes	B A C D	1.73333
Resolution	D C B A	0.66667

B. Intersubject Agreement

Before evaluating our metrics, we must demonstrate that subjects agree among themselves. Since there is no clear application of Fleiss' κ coefficient to measure agreement for these data, we express agreement by comparing our data to distributions representing agreement and disagreement, shown in Fig. 8.

- Perfect agreement: If subjects agreed completely with one another, they would all report the same ordering for a dimension. The average number of inversions across all subjects is 0 for this distribution. We can think of the average number of inversions as the amount of disorder in this distribution; in this case, there is no disorder because every subject agrees exactly with every other subject.
- Relative agreement: Given the subjective nature of how people perceive stories, we feel it may be impossible to achieve perfect agreement for a large group of subjects. It is more realistic to compare against a distribution which indicates high (but not perfect) agreement. We chose one such distribution (given in Fig. 8) which assumes that two thirds of the subjects will choose the most popular ordering, and then the function will decay exponentially by 3 from there. The average number of inversions across all subjects (i.e., the disorder) is 0.47 for this distribution.

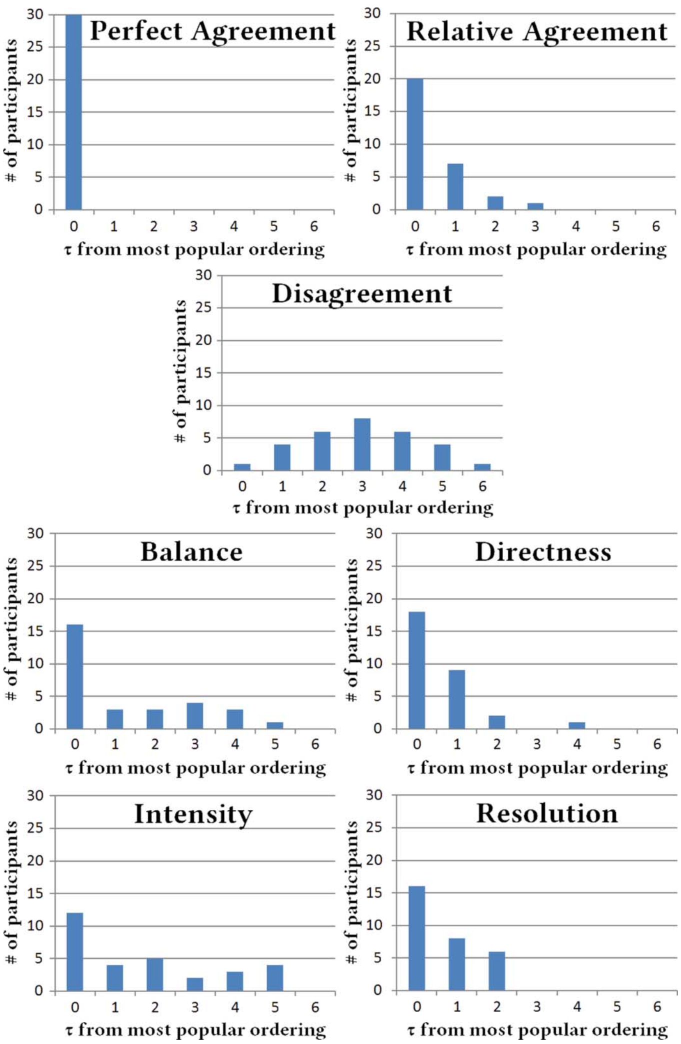


Fig. 8. Observed distributions for each dimension. These histograms show how many subjects (y-axis) chose an ordering that was some τ distance (x-axis) away from the most popular ordering for each dimension.

- Disagreement: If there is complete disagreement on the best ordering, we would expect answers to appear as if they were given at random. This would result in a uniform distribution across the 24 possible permutations for the four stories. That uniform distribution, when plotted as τ distance from the most popular ordering, is a roughly normal distribution (as seen in Fig. 8). The average number of inversions across all subjects (i.e., the disorder) is 3 for this distribution.

As a null hypothesis, we assume our observed distributions for each dimension will fit the disagreement distribution. We tested this using Fisher's exact test, which is similar to the χ^2 test but performs better for distributions with small expected values [42]. The p-values for these tests are

Dimension	P-Value
Balance	0.003
Directness	0.000
Stakes	0.028
Resolution	0.000

For all dimensions, $p < 0.05$, which means there is a statistically significant difference between our data and the disagreement distribution. The null hypothesis is rejected, that is, subjects do not disagree. Now we can evaluate the alternative hypothesis, that subjects agree on the most popular ordering. For this we employ the Bhattacharyya distance [43], which measures the distance between two discrete probability distributions. The Bhattacharyya distance is 0 when two distributions are the same, and approaches 1 as the distributions become less similar. Table VI gives the Bhattacharyya distances between our observed distributions and the perfect agreement, relative agreement, and disagreement distributions.

The dimensions for directness and resolution are more similar to the perfect agreement distribution than to the disagreement distribution; however, the dimensions of balance and stakes are more similar to disagreement than to perfect agreement. All four dimensions are most similar to relative agreement. These results support our hypothesis that subjects agree among themselves on a correct ordering for the four dimensions, especially for directness and resolution.

C. *Subject Agreement With Our Metrics*

For each dimension of conflict, Table VII presents the six orderings with the lowest τ_{avg} (the top six best orderings for that dimension according to subjects). Table VII also shows the ordering with the highest τ_{avg} (the worst ordering according to subjects) for each dimension. The orderings predicted by our formulas are highlighted in gray. For the dimensions of balance, directness, and resolution, the ordering predicted by our formula has the lowest τ_{avg}. For the dimension of stakes, the ordering predicated by our formula has the fifth lowest τ_{avg}, but this is only 0.6 inversions away from the most popular ordering. These data support our hypothesis that subjects will rank stories in the same order as our metrics.

D. *Discussion*

These results are promising, especially for balance, directness, and resolution. Several factors may account for what disagreement we did observe.

First, subjects may have misunderstood the dimension descriptions, which were intentionally brief and targeted at a high school reading level. We performed a pilot study before the experiment which indicated some confusion about these descriptions, especially stakes. We improved the descriptions with pilot study feedback. Subjects may also have misunderstood the stories themselves. At least one indicated a misunderstanding of the outcome of STORY D.

We assumed that each dimension could be measured independently of the others, but subjects may have perceived synergies between them. For example, if there was a lot on the line (high stakes) but there was little chance that the sorcerer would prevail (low balance), subjects might have given the story a low ranking for stakes. We hope to investigate how dimensions influence one another in future work.

Last, the two dimensions that showed the least subject agreement—balance and stakes—require the subject to measure them

TABLE VII
TOP SIX AND BOTTOM ONE ORDERINGS FOR DIMENSIONS BASED ON τ_{avg} (PREDICTED ORDERINGS IN GRAY)

Balance		Directness		Stakes		Resolution	
Order	τ_{avg}	Order	τ_{avg}	Order	τ_{avg}	Order	τ_{avg}
C D A B	1.26667	B A C D	0.56667	B A C D	1.73333	D C B A	0.66667
C D B A	1.66667	B A D C	0.96667	B A D C	1.93333	D C A B	1.20000
D C A B	1.73333	A B C D	1.36667	A B C D	2.13333	C D B A	1.40000
C A D B	2.00000	B C A D	1.36667	B C A D	2.26667	D B C A	1.40000
D C B A	2.13333	A B D C	1.76667	A B D C	2.33333	C D A B	1.93333
C B D A	2.26667	B D A C	1.90000	B D A C	2.33333	D A C B	1.93333
(17 omitted)	(17 omitted)	(17 omitted)	(17 omitted)	(17 omitted)	(17 omitted)	(17 omitted)	(17 omitted)
B A D C	4.73333	D C A B	5.43333	D C A B	4.26667	A B C D	5.33333

independently of the actual outcome of the story. If the protagonist appears likely to prevail, balance should be high regardless of whether he or she actually does succeed. At least two subjects had difficulty ignoring their knowledge of the actual outcome of the story. In future versions of this study, rather than ask subjects to ignore the ending, we intend to leave the ending out to avoid any bias introduced by their foreknowledge.

IX. STORY GENERATION WITH CPOCL

Having described a model of conflict and demonstrated that it operationalizes our definition, we now explore the expressive capabilities of the algorithm.

The CPOCL algorithm defines a space of solution plans containing conflict, and the seven dimensions of conflict provide a means of guiding the search toward different kinds of stories. Recall the simple problem and domain introduced in Fig. 1 from Section III. Three different solutions[2] are given in Fig. 9. Each one meets different authorial constraints. For brevity, we present only those conflict links involving Hank and measure the dimensions from Hank's point of view. All three stories begin the same way: Timmy is bitten by a snake, which prompts Hank to steal medicine from the general store.

STORY A is the shortest plan which also contains a conflict with a balance of 0.5 (recall that 0.5 is the "most balanced" from the author's point of view). This is due to the shootout between Hank and William, which is balanced because both parties are armed.

STORY B is the shortest plan which also contains a conflict with the highest possible stakes of 1. It is nearly identical to STORY A, but steps are ordered differently and different steps are marked as executed. These changes result in different kinds of conflict. When the shootout occurs before Hank makes it home, the stakes of the conflict with William increase because more is on the line, namely that Hank will not be able to heal Timmy.

STORY C is the shortest plan which also contains a conflict with the highest possible directness of 1. Hank manages to elude William only to get bitten by a snake himself. Now he is faced with an internal conflict: use the medicine to heal himself or to heal his son. A corresponding highly direct external conflict exists with Timmy for the same reason.

[2]After the story is planned, we assume that any nonexecuted steps which can be marked as executed are marked as such.

For this problem and domain, it is impossible for any conflict to have a resolution of 1 from Hank's point of view because a character cannot return to life after dying. Likewise, no conflict can have a resolution of -1. Even if both Hank and Timmy were bitten by snakes at the beginning, this would be two conflicts with resolutions of -0.6 and -0.4 for Hank, respectively.

X. FUTURE WORK

The larger context of this work is that of extending planning models to represent and reason about essential narrative phenomena. IPOCL extends POCL by modeling character intentionality, which enabled CPOCL to further extend that model to represent conflict. CPOCL is not necessarily intended as a finished story generation system in and of itself, but rather an important step forward for the larger project of plan-based story reasoning. This model does not yet address many other important narrative phenomena, so, in this section, we discuss some of the more important directions for future work.

A. Limitations of the Fabula Model

The example stories in Section IX make the limitations discussed in Sections VII-E and VIII-D clearer. The 11th and 12th conflict links in STORY C appear the same, but they are not. Hank's plan to heal Timmy threatens his plan to heal himself and his plan to heal himself threatens his plan to heal Timmy. The subplans mutually thwart each other, so it seems intuitive to group these conflicts together into one entity. But grouping conflict links is not always a straightforward process.

The fourth and fifth conflict links in STORY B also appear the same, but William's shootout threatens Hank's plan in two ways: he cannot travel back to the ranch if he is dead, and he also cannot heal Timmy if he is dead. The first experiment tells us that human readers are likely to report the latter conflict but not the former. We do not, however, expect readers to believe it is possible to travel home after being killed. A likely explanation for this disparity is that readers have grouped these conflict links conceptually.

Rather than consider conflicts at the fine-grained level of threatened causal links, it may be intuitive to reason only at the level of subplan pairs or character pairs. The difficulty here is deciding how to describe a group of conflict links which may all have different dimension values. A human with semantic knowledge about American Westerns may recognize which conflict links are the most important in a group, but there is

STORY A

Time	Step
1	**snakebite(Timmy)**
2	**travel(Hank,Ranch,GenStore)**
3	**take(Hank,Meds,Carl,GenStore)**
4	take(Carl,Meds,Hank,GenStore)
4	**travel(Hank,GenStore,Ranch)**
5	**heal(Hank,Timmy,Ranch)**
6	**travel(William,Saloon,Ranch)**
7	**shootout(William,Hank,Ranch)**

#	Participants	Reason	Dur.	Bal.	Dir.	Stk.	Res.
1	Hank vs. Fate	¬dying(Timmy)	1	0.95	0.00	0.60	-0.60
2	Hank vs. Carl	has(Carl,Meds)	2	0.80	0.50	0.60	0.60
3	Hank vs. William	alive(Hank)	4	0.50	0.50	0.40	-0.40
4	Hank vs. Carl	has(Hank,Meds)	1	0.83	0.00	0.60	0.60
5	Hank vs. Carl	has(Hank,Meds)	1	0.83	0.00	0.60	0.60
6	Hank vs. Carl	at(Hank,GenStore)	1	0.83	0.00	0.60	0.60

STORY B

Time	Step
1	**snakebite(Timmy)**
2	**travel(Hank,Ranch,GenStore)**
3	**take(Hank,Meds,Carl,GenStore)**
4	**travel(William,Saloon,GenStore)**
5	**shootout(William,Hank,GenStore)**
6	travel(Hank,GenStore,Ranch)
6	heal(Hank,Timmy,Ranch)
6	**take(Carl,Meds,Hank,GenStore)**

#	Participants	Reason	Dur.	Bal.	Dir.	Stk.	Res.
1	Hank vs. Fate	¬dying(Timmy)	1	0.95	0.00	0.60	-0.60
2	Hank vs. Carl	has(Carl,Meds)	2	0.80	0.50	0.60	0.60
3	Hank vs. William	alive(Hank)	1	0.50	0.00	0.40	0.00
4	Hank vs. William	alive(Hank)	2	0.67	0.50	1.00	-0.40
5	Hank vs. William	alive(Hank)	2	0.67	0.50	1.00	-0.40
6	Hank vs. William	at(Hank,GenStore)	2	0.67	0.50	1.00	-0.40
7	Hank vs. Carl	has(Hank,Meds)	2	0.83	0.50	0.60	0.00
8	Hank vs. Carl	has(Hank,Meds)	2	0.83	0.50	0.60	0.00
9	Hank vs. Carl	at(Hank,GenStore)	2	0.83	0.50	0.60	0.00

STORY C

Time	Step
1	**snakebite(Timmy)**
2	**travel(Hank,Ranch,GenStore)**
3	**take(Hank,Meds,Carl,GenStore)**
4	travel(William,Saloon,GenStore)
4	shootout(William,Hank,GenStore)
4	take(Carl,Meds,Hank,GenStore)
4	**travel(Hank,GenStore,Ranch)**
5	**snakebite(Hank)**
6	**heal(Hank,Timmy,Ranch)**
7	heal(Hank,Hank,Ranch)

#	Participants	Reason	Dur.	Bal.	Dir.	Stk.	Res.
1	Hank vs. Fate	¬dying(Timmy)	1	0.95	0.00	0.60	-0.60
2	Hank vs. Fate	¬dying(Hank)	5	0.95	0.00	0.40	-0.40
3	Hank vs. Carl	has(Carl,Meds)	2	0.80	0.50	0.60	0.60
4	Hank vs. William	alive(Hank)	1	0.50	0.00	0.40	0.00
5	Hank vs. William	alive(Hank)	1	0.67	0.00	1.00	0.60
6	Hank vs. William	alive(Hank)	1	0.67	0.00	1.00	0.60
7	Hank vs. William	at(Hank,GenStore)	1	0.67	0.00	1.00	0.60
8	Hank vs. Carl	has(Hank,Meds)	2	0.83	0.00	0.60	0.60
9	Hank vs. Carl	has(Hank,Meds)	2	0.83	0.00	0.60	0.60
10	Hank vs. Carl	at(Hank,GenStore)	2	0.83	0.00	0.60	0.60
11	Hank vs. Hank	has(Hank,Meds)	1	0.50	1.00	0.20	0.60
12	Hank vs. Hank	has(Hank,Meds)	1	0.50	1.00	0.20	0.60
13	Hank vs. Timmy	has(Hank,Meds)	1	0.50	1.00	0.20	0.60
14	Hank vs. Timmy	has(Hank,Meds)	1	0.50	1.00	0.20	0.60

Fig. 9. Three Western stories generated by CPOCL for the problem and domain in Fig. 1. Bold steps are executed; nonbold steps are nonexecuted. Next to each story is a table showing all conflicts involving Hank along with the participants, reason, duration (Dur.), balance (Bal.), directness (Dir.), stakes (Stk.), and resolution (Res.) from Hank's point of view.

nothing inherent in the structure of a CPOCL plan to signify this.

B. Reasoning About Discourse

Currently the CPOCL model is concerned only with fabula, so the most important direction for future work will be the ability to reason about conflict at the discourse level. Discourse techniques (such as telling a story out of order, deceiving the audience, etc.) can have a significant impact on which conflicts are salient to the audience and on perceived dimension values.

The first experiment demonstrates that the structure of a story can indicate a superset of conflicts that readers actually perceive. We suspect that psychological research on how an audience attends narrative content during the reading process [44] will provide a foundation for modeling which conflicts, of the many defined by CPOCL, are most apparent to the reader. Initial work on this is already underway [11].

It will also be interesting to investigate how imperfect information affects the perception of the dimensions. In *Star Wars: The Empire Strikes Back*, the audience learns that Darth Vader is Luke Skywalker's father, and this new information changes their perception of the directness of the conflict between Luke and Vader. This fact was always true at the fabula level, but withholding this information until later in the discourse allows the filmmaker to achieve a different effect on the audience.

C. Planning Heuristics

Regarding the algorithm, the most significant limitation is the difficulty inherent in providing effective heuristic guidance to the planner during search. All planners work by building partial plans into complete plans (though what constitutes a "partial plan" differs between algorithms). Conflict as we have defined it is a property of sequences of actions, and only until both sequences are fully constructed can the dimensions of that conflict be measured. Thus, it is difficult to evaluate the dimension values of conflicts in a partial plan. This is true both of POCL-style planners and of state–space planners, which have a different but equally problematic set of unknowns when building a partial plan [45].

XI. Conclusion

Conflict is an essential phenomenon in narrative. Previous work in plan-based models of narrative have focused on producing coherent stories in which characters act intentionally. The CPOCL model and algorithm extend this work based on narratology research to allow for conflict by reasoning about threatened causal links that arise between subplans. CPOCL also defines seven dimensions of conflict which aid in story analysis and can be used by authors to guide the output of a story generator.

The CPOCL model and dimensions have been empirically evaluated to demonstrate that they correctly operationalize our chosen narratological definition. These experiments also reveal the limitations of CPOCL and the need for further work, especially in the area of semantic understanding and discourse reasoning. In conclusion, we believe this work represents progress toward the goal of empowering computer systems to automatically create and adapt plots based on the appealing structural properties that conflict provides.

References

[1] H. P. Abbott, *The Cambridge Introduction to Narrative*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[2] D. Herman, M. Jahn, and M. L. Ryan, *Routledge Encyclopedia of Narrative Theory*. Oxon, U.K.: Routledge, 2005.

[3] C. Brooks and R. P. Warren, *Understanding Fiction*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.

[4] R. J. Gerrig, *Experiencing Narrative Worlds: On the Psychological Activities of Reading*. New Haven, CT, USA: Yale Univ. Press, 1993.

[5] N. Carroll, *Theorizing the Moving Image*. Cambridge, U.K.: Cambridge Univ. Press, 1996.

[6] M. Tierno, *Aristotle's Poetics for Screenwriters: Storytelling Secrets From the Greatest Mind in Western Civilization*. New York, NY, USA: Hyperion, 2002.

[7] R. M. Young, "Notes on the use of plan structures in the creation of interactive plot," in *Proc. Assoc. Adv. Artif. Intell. Fall Symp. Narrative Intell.*, 1999, pp. 164–167.

[8] M. Bal, *Narratology: Introduction to the Theory of Narrative*. Toronto, ON, Canada: Univ. Toronto Press, 1997.

[9] B. Bae and R. Young, "A use of flashback and foreshadowing for surprise arousal in narrative using a plan-based approach," in *Proc. Int. Conf. Interactive Digital Storytelling*, 2008, pp. 156–167.

[10] J. M. Thomas and R. M. Young, "Using task-based modeling to generate scaffolding in narrative-guided exploratory learning environments," in *Proc. 14th Int. Conf. Artif. Intell. Edu.*, 2009, pp. 107–114.

[11] R. E. Cardona-Rivera, B. A. Cassell, S. G. Ware, and R. M. Young, "Indexter: A computational model of the event-indexing situation model for characterizing narratives," in *Proc. Comput. Models Narrative Workshop*, 2012, pp. 34–43.

[12] M. O. Riedl and R. M. Young, "Narrative planning: Balancing plot and character," *J. Artif. Intell. Res.*, vol. 39, pp. 217–268, 2010.

[13] L. Egri, *The Art of Dramatic Writing*. Rockville, MD, USA: Wildside, 1988.

[14] A. Dibell, *Elements of Fiction Writing: Plot*. Cincinnati, OH, USA: Writer's Digest, 1988.

[15] M. Bratman, *Intentions, Plans, and Practical Reason*. Cambridge, MA, USA: Harvard Univ. Press, 1987.

[16] N. Szilas, "IDtension: A narrative engine for interactive drama," in *Proc. 1st Conf. Technol. Interactive Digital Storytelling Entertain.*, 2003, vol. 3, pp. 187–203.

[17] J. R. Meehan, "Tale-spin, an interactive program that writes stories," in *Proc. 5th Int. Joint Conf. Artif. Intell.*, 1977, pp. 91–98.

[18] D. Thue, V. Bulitko, and M. Spetch, "Making stories player-specific: Delayed authoring in interactive storytelling," in *Proc. 1st Int. Conf. Interactive Digital Storytelling*, 2008, pp. 230–241.

[19] M. Lebowitz, "Story-telling as planning and learning," *Poetics*, vol. 14, no. 6, pp. 483–502, 1985.

[20] R. P. É. Ý. Pérez and M. Sharples, "MEXICA: A computer model of a cognitive account of creative writing," *J. Exp. Theor. AI*, vol. 13, no. 2, pp. 119–139, 2001.

[21] T. C. Smith and I. H. Witten, "A planning mechanism for generating story text," *Literary Linguistic Comput.*, vol. 2, no. 2, pp. 119–126, 1987.

[22] H. Barber and D. Kudenko, "Dynamic generation of dilemma-based interactive narratives," in *Proc. 3rd AAAI Conf. Artif. Intell. Interactive Digital Entertain.*, 2007, pp. 2–7.

[23] R. Aylett, M. Vala, P. Sequeira, and A. Paiva, "Fearnot!—An emergent narrative approach to virtual dramas for anti-bullying education," in *Proc. 4th Int. Conf. Virtual Storytelling, Using Virtual Reality Technol. Storytelling*, 2007, pp. 202–205.

[24] Y.-G. Cheong, R. Khaled, C. Grappiolo, J. Campos, C. Martinho, G. P. Ingram, A. Paiva, and G. Yannakakis, "A computational approach towards conflict resolution for serious games," in *Proc. 6th Int. Conf. Found. Digital Games*, 2011, pp. 15–22.

[25] J. Carbonell, "Counterplanning: A strategy-based model of adversary planning in real-world situations," *Artif. Intell.*, vol. 16, no. 3, pp. 295–329, 1981.

[26] J. Gratch, "Émile: Marshalling passions in training and education," in *Proc. 4th Int. Conf. Autonom. Agents*, 2000, pp. 325–332.

[27] S. Marsella and J. Gratch, "Ema: A process model of appraisal dynamics," *Cogn. Syst. Res.*, vol. 10, no. 1, pp. 70–90, 2009.

[28] C. Battaglino and R. Damiano, "Emotional appraisal of moral dilemma in characters," in *Proc. 5th Int. Conf. Interactive Storytelling*, 2012, pp. 150–161.

[29] G. N. Yannakakis, "How to model and augment player satisfaction: A review," presented at the 1st Workshop Child Comput. Interaction, Chania, Crete, Greece, Oct. 20–22, 2008.

[30] F. Peinado and P. Gervás, "Evaluation of automatic generation of basic stories," *New Generation Comput.*, vol. 24, no. 3, pp. 289–302, 2006.

[31] F. Zambetta, A. Nash, and P. Smith, "Two families: Dynamical policy models in interactive storytelling," in *Proc. Australasian Conf. Interactive Entertain.*, 2007, pp. 1–8.

[32] M. Booth, "The AI systems of left 4 dead," presented at the Artif. Intell. Interactive Digital Storytelling, 2009, keynote.

[33] M. Williamson and S. Hanks, "Flaw selection strategies for value-directed planning," in *Proc. 3rd Int. Conf. Artif. Intell. Planning Syst.*, 1996, vol. 23, pp. 244–252.

[34] D. S. Weld, "An introduction to least commitment planning," *AI Mag.*, vol. 15, no. 4, pp. 27–61, 1994.

[35] D. Herman, *Story Logic*. Lincoln, NE, USA: Univ. Nebraska Press, 2002.

[36] S. G. Ware, R. M. Young, B. Harrison, and D. L. Roberts, "Four quantitative metrics describing narrative conflict," in *Proc. 5th Int. Conf. Interactive Digital Storytelling*, 2012, pp. 18–29.

[37] S. Kambhampati, C. A. Knoblock, and Q. Yang, "Planning as refinement search: A unified framework for evaluating design tradeoffs in partial-order planning," *Artif. Intell.*, vol. 76, no. 1–2, pp. 167–238, 1995.

[38] C. Crawford, *Chris Crawford on Game Design*. New York, NY, USA: New Riders, 2003.

[39] S. G. Ware and R. M. Young, "Validating a plan-based model of narrative conflict," in *Proc. 7th Int. Conf. Found. Digital Games*, 2012, pp. 220–227.

[40] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, pp. 159–174, 1977.

[41] M. G. Kendall, *Rank Correlation Methods*. Duxbury, MA, USA: Griffin, 1948.

[42] J. L. Fleiss, B. Levin, and M. C. Paik, *Statistical Methods for Rates and Proportions*, 3rd ed. New York, NY, USA: Wiley, 2003.

[43] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, vol. 35, no. 99–109, p. 4, 1943.

[44] R. A. Zwaan and G. A. Radvansky, "Situation models in language comprehension and memory," *Psychol. Bull.*, vol. 123, no. 2, pp. 162–185, 1998.

[45] S. G. Ware, "The intentional fast-forward narrative planner," in *Proc. 5th Intell. Narrative Technol. Workshop*, 2012, pp. 57–62.

Stephen G. Ware (S'12–M'13) received the B.S. degree in computer science and philosophy from Loyola University, New Orleans, LA, USA, in 2008 and the M.S. and Ph.D. degrees in computer science from North Carolina State University, Raleigh, NC, USA, in 2011 and 2014, respectively.

He is currently an Assistant Professor of Computer Science at the University of New Orleans, New Orleans, LA, USA. His research interests are in computer narrative, especially automated planning, knowledge representation, narratology, cognitive science, and artificial intelligence in games.

Dr. Ware is a member of the Association for Computing Machinery and the Association for the Advancement of Artificial Intelligence.

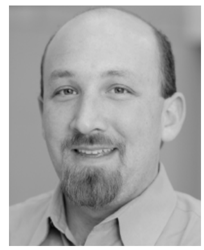
R. Michael Young (M'98–SM'07) received the B.S. degree in computer science from the California State University at Sacramento, Sacramento, CA, USA, in 1984, the M.S. degree in computer science with a concentration in symbolic systems from Stanford University, Stanford, CA, USA, in 1988, and the Ph.D. degree in intelligent systems from the University of Pittsburgh, Pittsburgh, PA, USA, in 1998.

He worked as a Postdoctoral Fellow at the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA. He is currently an Associate Professor of Computer Science and Executive Director of the Digital Games Research Center, North Carolina State University, Raleigh, NC, USA. His research interests span the computational modeling of narrative, the use of artificial intelligence techniques within computer games, and virtual worlds.

Dr. Young is a senior member of the Association for Computing Machinery and a member of the Association for the Advancement of Artificial Intelligence and of the International Game Developers Association.

Brent Harrison (S'13) received the B.S. degree in computer science and the B.A. degree in English from Auburn University, Auburn, AL, USA, in 2008 and the M.S. and Ph.D. degrees in computer science from North Carolina State University, Raleigh, NC, USA, in 2012 and 2014, respectively.

He is currently a Postdoctoral Fellow at the Georgia Institute of Technology, Atlanta, GA, USA. His research interests are in machine learning, player modeling, and artificial intelligence for adaptive games.



David L. Roberts received the B.A. degree in computer science and mathematics from Colgate University, Hamilton, NY, USA, in 2003 and the Ph.D. degree in computer science from the College of Computing, Georgia Institute of Technology, Atlanta, GA, USA, in 2010.

He is currently an Assistant Professor of Computer Science at North Carolina State University, Raleigh, NC, USA. His research interests lie at the intersection of machine learning, social and behavioral psychology, and human–computer interaction. He has a particular focus on computation as a tool to provide insight into human behavior in narrative, virtual world, and game environments.

Dr. Roberts is a member of the Association for the Advancement of Artificial Intelligence and the Association for Computing Machinery.