

Computational Subjectivity in Virtual World Avatars

Michael Mateas

Carnegie Mellon University
Computer Science Department
Pittsburgh, PA 15213
michaelm@cs.cmu.edu

Abstract

This paper describes subjective avatars, a kind of socially intelligent agent. Such avatars, employed in story worlds, have autonomous emotional reactions to events in the world and keep track of story context. Much of this internal state is concerned with the avatar's social relationship with other characters in the world. The emotional state and story context are used to provide subjective descriptions of sensory information and to generate a stream of thought. The purpose of such description is to help a user gain a deeper understanding of the role they are playing in the story world.

Introduction

The goal of the Oz project (Bates 1992) at CMU is to build dramatically interesting virtual worlds inhabited by believable agents - autonomous characters exhibiting rich personalities, emotions and social interactions. In many of these worlds, the player is herself a character in the story, experiencing the world from a first person perspective. Typically, the player's representation within the world - her avatar - is passive. The avatar performs actions as fully specified by the player, and reports events (reporting events can mean rendering a 3D scene or generating descriptive text) in a pseudo-objective manner (*pseudo-objective* because any description encodes the bias of the world author). This paper describes an alternative: a subjective avatar with autonomous interpretations of the world.

Why Subjective Avatars?

I want the user to step into the shoes of a character, experiencing a story from an alien perspective. In this manner the user gains an empathic understanding of a character by *being* this character. In non-interactive drama (movies, theater), an audience is able to gain insights into the subjective experience of characters precisely because the experience is non-interactive; the characters in the drama make decisions different than those that audience members might make. For example, consider the movie *The Remains of the Day*:

A rule bound head butler's world of manners and decorum in the household he maintains is tested by the

arrival of a housekeeper who falls in love with him in post-W.W.I Britain. The possibility of romance and his master's cultivation of ties with the Nazis challenge his carefully maintained veneer of servitude. (Loh 1993)

As an audience, we are able to gain empathy (subjective insight) into the butler because we can watch the butler make decisions within the context of his personal history and social role. However, in a *Remains of the Day* interactive world, how would a user know how to act like this W.W.I era English butler? What is to keep the user from immediately standing up to the master, thus derailing the story and preventing any empathic insight into this situation? The hope is that if the user's avatar filters and interprets the world in a manner consistent with this W.W.I era butler, the user will begin to feel like this butler, gaining a deeper understanding of the message the author wants to convey.

In addition, if the user really begins to *feel* like this butler, she will implicitly constrain her actions; for example, she won't punch the master in the nose. Such constraints are important: no matter how sophisticated the characters and plot control system might be in some story world, the user must be willing to "play along" in order to have a high quality story experience (this is the equivalent of suspending disbelief in a play or movie). However, such manipulations of the user's freedom must remain implicit in order to maintain the special vividness interaction lends to a story (Kelso, Weyhrauch, and Bates 1992).

Once the avatar is actively interpreting the world, the possibility is open for employing abstract, symbolic or surreal descriptions. Some of these possibilities are explored in (Smith and Bates 1989). Thus, in addition to improving the user's experience, subjective avatars open up new artistic possibilities in the construction of dramatic worlds.

Fastfood World

I'm currently experimenting with subjective avatars within the Oz text-based system. The specific world I've built within this framework is *Fastfood World*, a McJob (Howe and Strauss 1993) from hell scenario set in a fastfood restaurant.

The Framework

The text-based system provides a framework for building worlds which are rendered in a manner similar to text-based adventure games or multi-user dungeons. This framework includes a non-real-time version of Tok, the Oz agent architecture (Bates, Loyall, and Reilly 1992), an object oriented physical world ontology for building the physical world simulation, a sense ontology for modeling the propagation of sensory information (sensory information can include descriptions of objects, actions and events), and a natural language generator for rendering into text descriptions of objects, actions and events (Kantrowitz and Bates 1992). Note: the text spoken by characters is canned; it is not generated from semantic primitives.

The Characters

In Fastfood World, the user plays a young worker in his early 20s. The worker is frustrated by the dead-end job he finds himself in. He is interested in richly exploring life and thinking about the Big Questions but feels dragged down by his current job, which leaves him too depressed and apathetic to passionately explore his own interests.

Just a quick note on terminology: for the rest of this paper I will refer to the character of the young worker as the player. The player is the user in the role of the young worker. I will also use the masculine pronoun for this character, since, though the user may be of either gender, the role being played is a masculine role.

The player's nemesis is the manager, Barry. Barry, now in his early 30s, has been working at the restaurant since he was 16. He's proud of having worked his way up from clerk to assistant manager. He uses every interaction to assert his superiority over the player. He yells at the player in front of customers, gives long-winded explanations to simple questions, and reminds the player of past screw-ups. A common theme in the harangues he directs at the player is the need to stop dreaming, to stop thinking about the big picture and pay attention to details.

The player also interacts with a customer named Celia. Celia is a garrulous old lady who is more interested in social interaction than in ordering her food. She asks the worker questions about the food (e.g. "How greasy are your fries?"), and tends to break off into reminiscences. A common theme in her conversation is how the small concerns of day-to-day life tend to fade away as one grows older; looking back on her long life, what was important (and remembered) are themes, feelings, relationships.

The Story

The player's experience in this world is divided into three parts. First, the player comes into work late. He's concerned about being caught. If he can only get to the register and sign in before Barry sees him, he might be OK. At this point, Barry is wandering about the restaurant, glancing at his watch and looking for the worker. If the player starts working on the register before Barry chews him out, Barry makes a mental note to chew the player out

later. If Barry catches the player before he's started working, Barry yells at him for being late.

Second, after the player has either escaped being yelled at or not, the player will try to sign into the register. To his horror, the player discovers that he is not able to sign into the register because a previous worker left an incomplete transaction on it. The only person who can clear the register is Barry. The player knows, however, that if he asks Barry, Barry will use this as an opportunity to assert his superiority over the player. If the player continues trying to sign into the register and Barry notices this, he chews the player out for not asking for help (e.g. "we've all got to be a team"), and then "helps" the player. If the player asks for help, Barry chews the player out for not having learned to operate the register yet (e.g. "how many times do I have to show you?") and then "helps" the player. Barry's help consists of a long winded explanation of how you sign into the register (even though the player already knows how to sign in and exactly what the problem is) interspersed with descriptions of Barry's own rise to the glory of assistant manager. It is during this interaction that Barry stresses the importance of focusing on details (a petty focus on details), thus revealing one pole of the tension the player will have to resolve. Eventually Barry clears the register transaction. Incidentally, there is no way for the player to clear the register themselves. Interactive drama is not about solving puzzles; it is about relationships with characters. In this particular case, the player is in a "damned if you do, damned if you don't" situation in which there is no way to avoid having an unpleasant interaction with Barry.

Finally, Celia comes in to order food. The player first perceives her as a batty old woman. Celia seems disorganized and confused. She's more interested in conversation than in ordering food. During her conversation she begins revealing how she's forgotten many of the mundane details in her life; what remains as important are themes, feelings, and relationships. She thus reveals the second pole of the tension facing the player (petty details vs. meaningful experience). During this time, Barry comes out of his office repeatedly, admonishing the player to keep the line moving. If the player succumbs to his fear of Barry and the general bad mood that his job puts him in, the player will continue to perceive Celia as batty and irritating and Barry as dominant and scary. Eventually Celia will finish ordering and the story experience ends (petty details won). However, there is an opportunity for a transformative experience. If the player begins interacting with Celia more deeply, he begins perceiving her as first interesting, and eventually, wise. The player begins perceiving Barry as less scary, and eventually sad and pathetic. The story experience ends with Barry's psychological hold on the player broken (meaningful experience won).

Subjective State

Now that I've described the general motivation for subjective avatars and the specific story within which I'm

exploring the concept, it's time to describe how a subjective interpretation of the world can be implemented.

In order for an avatar to provide a subjective interpretation for the player, it must both respond to events in the world by maintaining some sort of state and communicate this state to the player. First I'll describe the mechanisms by which the avatar maintains subjective state. Currently, the avatar's subjective state consists of emotional state (emotional responses to events) and story context.

Emotional State

To maintain emotional state, I make use of Em, the Oz model of emotion (Neil Reilly 1996). In Em, emotions are generated primarily in response to goal processing events and attitudes. To enable the generation of goal processing emotions, the author of an agent provides annotations for agent's goals such as importance, likelihood-of-success, likelihood-of-failure, failure-responsibility, etc. These annotations allow Em to generate emotions as goals are created, as beliefs change about the likelihood of goals succeeding or failing, and as goals actually succeed or fail. For example, if a goal with non-zero importance fails, distress will be generated. If the goal was annotated with a failure-responsibility function, this function will be called to infer the cause of the failure; anger towards the agent who caused the failure will be generated. If, prior to the failure, there was a non-zero belief that the goal would succeed, disappointment will also be generated. In general, at any given moment, an agent's emotional state will contain several emotions with non-zero values. Over time, Em decays emotions.

In order for the avatar to have goal processing emotions, it must be processing some goals. But, since the avatar never engages in autonomous behaviors that directly result in action (the avatar never moves it's own muscles), what kinds of goals can an avatar have? An avatar can have passive goals, that is, a goal for some event to happen in the world but for which you don't directly take action. Such goals passively wait for some event to occur in the world in order to succeed or fail.

As an example of goal-based emotion processing, consider the avatar goal to avoid being chewed out for being late. The avatar's internal representation of this goal follows.

```
(with
  (ignore-failure 'except-em
  (importance 4)
  (compute-lof
    (lambda () (player-be-chewed-out-lof)))
  (failure-responsibility
    (lambda () (list $$barry)))
  (subgoal avoid-being-chewed-out-for-late))
```

Some behavior is calling the subgoal avoid-being-chewed-out-for-late. If it fails, emotions are generated but the failure does not cause the enclosing behavior to fail. The

importance of this goal is 4 on a scale of 1 to 10. Some function called player-be-chewed-out-lof determines the avatar's belief in the likelihood of failure of this goal. If the goal fails, the agent responsible is Barry.

The avatar's internal representation of the behavior for avoid-being-chewed-out-for-late follows.

```
(parallel-production
  avoid-being-chewed-out-for-late ()
  (subgoal
    succeed-in-avoiding-chewing-out-for-late)
  (with
    (priority-modifier 1)
    effect-only-no-fail
    (subgoal
      fail-in-avoiding-chewing-out-for-late)))
```

Avoid-being-chewed-out-for-late simultaneously (in parallel) examines the world for evidence that the player got to work before Barry could chew him out and for evidence that Barry *is* chewing him out. Both the succeed-in... and fail-in... subgoals are passive goals. They sit and wait for their appropriate conditions to be met. While those two goals are waiting, the avatar can pursue other goals. If the succeed-in... goal succeeds, then avoid-being-chewed-out-for-late succeeds. If the fail-in... subgoal fails, then avoid-being-chewed-out-for-late fails. In either case, the appropriate emotions (joy, or anger and distress) are generated.

In addition to goal processing emotions, the avatar makes use of attitude-based emotions. Attitudes record the fact that the avatar feels a certain way towards certain objects. Every time the avatar senses the appropriate object in its vicinity, an emotion is generated. For example, the avatar fears Barry with a strength of 3 (1 to 10 scale). Every time Barry is in the same location as the avatar, a fear emotion of strength 3 towards Barry is generated. This is in addition to any goal processing emotions the avatar might be having.

Story Context

In addition to emotion processing, the avatar keeps track of where it is in the story (or, at least, the avatar's conception of the story). This is done so as to organize the avatar's goals and simplify the writing of behaviors. At any given moment, there are a set of goals active in the avatar. Some of these are passive goals for emotion generation; others describe specific objects or events to the user or report the avatar's thoughts to the user (described below). The behaviors associated with these goals are watching for certain events or sequences of events to occur in the world. Depending on where the player is in the story experience, the same event may need to trigger different reactions. Alternatively, reacting to a given event may only be appropriate during certain parts of the story. In this case, even if the trigger event never occurs, the goal should be removed from the set of active goals when the story has passed a given point without requiring that the goal succeed

or fail. Explicitly maintaining a story context pushes the context information into the set of active goals instead of requiring this information to be included in the test expression of every behavior.

For example, the following behavior represents the avatar's most abstract understanding of the structure of the story.

```
(sequential-production a-day-in-the-store ()
  (subgoal come-into-work-late)
  (subgoal setup-register)
  (subgoal wait-on-customer))
```

A day in the fastfood restaurant consists of three sequential segments: coming into work late, setting up the register, and waiting on a customer. The behaviors associated with each of these subgoals watch for indications that their context has passed (e.g. "I'm done coming into work late."). The goal then succeeds and a behavior associated with the next goal is chosen.

The following behavior for the goal come-into-work-late illustrates how an active behavior defines a story context.

```
(parallel-production come-into-work-late ()
  (with
    (ignore-failure 'except-em)
    (importance 4)
    (compute-lof
      (lambda () (player-be-chewed-out-lof)))
    (failure-responsibility
      (lambda () (list $$barry)))
    (subgoal avoid-being-chewed-out-for-late))
  (with
    effect-only
    (subgoal think-about-being-late)))
```

The first subgoal, avoid-being-chewed-out-for-late was described above. It's job is to generate either positive or negative emotions depending on whether the player is chewed out or not. The second subgoal generates a stream of thought which is reported to the user under appropriate circumstances (described below). The main point here is that this particular opportunity for emotion generation and this particular stream of thought are only appropriate when the player first comes into work. The (ignore-failure 'except-em) and effect-only annotations ensure that the generation of a stream of thought has no effect on terminating this story context, while either the success *or* failure of avoid-being-chewed-out-for-late causes this behavior to terminate with success, resulting in a transition to the next story context.

Narrative Effects

Once the avatar is maintaining a subjective state, that state must influence the user's experience. That is, the subjective state must somehow be expressed. So far I've experimented with two classes of effects: manipulating sensory

descriptions and generating a stream of thought.

Sensory Descriptions

The default Oz avatar produces pseudo-objective factual descriptions of the world. The following is a short trace of the default avatar in Fastfood World.

```
You are in the counter area.
... <some stuff deleted> ...
The deep fat fryer, the cash register and the
food tray are in the counter area.
The three hamburgers are on the food tray.
You are wearing your clothes.
You hear the hot oil sizzling in the fryer.
```

```
PLAYER> wait
... <a turn deleted> ...
```

```
You wait.
You hear the hot oil sizzling in the fryer.
Barry is speaking to you.
Barry's voice says "Wait a minute there, buster"
```

```
PLAYER> look at hamburgers
```

```
You look at the hamburgers.
The three hamburgers are on the food tray.
You hear the hot oil sizzling in the fryer.
Barry goes to the counter area.
Barry is no longer in the window area.
```

Each turn, sensory information propagates through the physical world graph. Each sensory datum describes an object, action or event. The default avatar constructs intermediate semantic representations called groups to describe the sensory data received during a given clock tick. These groups are then passed to the natural language generator.

In the current implementation, subjective descriptions are generated by intercepting the groups before they are sent to the generator. The groups are passed through a set of rules that, based on the structure of the group and the current emotional state, render the group in natural language. Currently, rules generate descriptions of groups using canned text. If no rule matches, the group is eventually passed to the natural language generator.

The typical structure of such a rule is as follows.

```
(sequential-production describe-something (group)
  (precondition
    (and (<group has a given form>)
         (<avatar in a given state>)))
  (<render the group in a given way>))
```

Here is a short trace of the subjective avatar in the Fastfood World.

```
You are in the counter area.
```

... <some stuff deleted> ...
The deep fat fryer, the cash register and the food tray are in the counter area.
The three hamburgers are on the food tray.
You are wearing your clothes.
You hear the hot oil sizzling in the fryer.

PLAYER> wait
... <a turn deleted> ...

You wait.
With a vindictive gleam in his eye, Barry snaps
"Wait a minute there, buster"

PLAYER> look at hamburgers

You look at the hamburgers.
The faceless crowd of hamburgers sits on the food tray.
As if pop rocks had been poured directly on your brain, the hideous sizzle of hot grease emanates from the fryer.
Barry marches toward you from the drive-up window station.

In this trace, several effects are present. In the first turn, the player hears the sizzle of hot grease. In the second turn, this is absent. This occurs because one of the rules checks if this is the first time the player has been behind the counter where he can hear the fryer. If so, the group describing the fryer's sound is passed onto the natural language generator. After the sound has been initially described (thus making the player aware that the fryer does make noise), the sound is not described unless the player is in a particularly fearful state.

In the second turn, the player is not feeling enough fear, so the description of the fryer sound is bypassed. However, the description of Barry speaking to (snapping at) the player is altered. This is due to a moderate amount of fear generated by an increase in the likelihood of failure (the player's belief) of the avoid-being-chewed-out goal failing. This increase in the likelihood of failure is due to hearing Barry speak while the player is still coming in late to work (the player hasn't started working before Barry noticed him).

In the third turn, the description of the hamburgers has changed since the description in the first turn. This is due to a rule that renders any group describing more than two similar objects as "a faceless mass" if the player is in a bad enough mood. Since the matching criterion on the group is fairly general (any description of a multiple number of objects), I hope that such a narrative effect can become a repeated device, a way of communicating a sense of depressing conformity by repeating a key phrase in a variety of contexts. The ability to match on a generic emotion (such as a bad mood) rather than just on specific emotions (e.g. anger at Barry for some specific reason) is facilitated by combination rules in Em that can combine the types and intensities of more specific emotions into an

intensity on a more general emotion.

The sizzle of hot grease is back, but this time in a more colorful form. Barry moving into the counter area (same physical space with the player) initiated a new burst of fear from a fear attitude the player holds towards Barry. This new burst of fear was enough to activate the rule that describes the sound of the fryer.

Finally, Barry's movement into the counter area is described by one of a set of rules that watch for any movement Barry makes. Barry's movements are described differently depending on the current emotional state.

Stream of Thought

In addition to manipulating sensory descriptions, I've also experimented with generating a stream of thought. I hope to provide the user with additional orientation to the character they are playing by providing access to the inner life of this character.

The first category of thoughts in this stream of thought is commentary on the current situation. As an example, in the behavior come-into-work-late (described above), one of the subgoals is think-about-being-late. When the user first enters the world, this behavior produces "Damn! I'm late for work. I hope Barry doesn't notice me." If, for several moves in a row, the player makes no progress towards getting to the counter, it then produces "If I can only get to the register and sign in before Barry sees me." When the player first arrives in the counter area, and if the message above wasn't produced (the player didn't take awhile to get to the counter), the behavior produces "If I can just get signed into the register before Barry sees me, I'll be all right." In this example, the commentary takes the form of hints, helping the user to understand the current story situation. In general, such commentary can be less direct, such as thinking "Yeah, right!" when Barry says something like "You need to be a team player."

The recall of memories is a mechanism that could help a user to understand their role more deeply over time. Fragments of events which occurred prior to the story experience could be recalled in appropriate story and emotional contexts. This would give the user access to the prior life of the avatar. I have not yet implemented such memories.

Daydreams may be another mechanism for communicating role information to the user. Mueller's work on daydreaming (Mueller 1990) is particularly relevant. I have not yet implemented any daydreaming.

What's Actually Been Built

The full Fastfood World story described in the beginning of the paper has not yet been built. As of the time of this writing, Barry's behaviors for dealing with a late worker and helping a worker with the register (including his long harangues) have been written. Celia's behaviors for ordering food (but not for reminiscing about her life) have been written. In the avatar, all the behaviors for coming

into work late, as well as a number of description rules which apply over the entire experience have been written.

Related Work

Hayes-Roth's work on improvisational puppets (Hayes-Roth and van Gent 1996) shares a similar goal of helping a user to play a role. Her work, however, focuses on *actions* performed by an avatar rather than on maintaining a complex internal state. In her work, she provides the user with an interface for giving the avatar high level directives. The avatar then translates these directives into concrete actions consistent with the character. In my work, the space of actions the avatar can engage in is not explicitly presented to the user. Instead, I hope to implicitly guide the user in their role by providing access to subjective sensory descriptions and a stream of thought.

Galyean's work (Galyean 1995) employs narrative manipulation of point of view in an immersive, graphical story world. This manipulation guides the user through the story. The narrative manipulation, however, is controlled by a central story manager rather than by the internal state of an avatar. His system also provides no facilities for access to a stream of thought.

Conclusion

In this paper I've described a kind of socially intelligent agent: subjective avatars. Such avatars, employed in story worlds, have autonomous emotional reactions to events in the world and keep track of story context. Much of this context is concerned with the avatar's social relationship with other characters in the world. The emotional state and story context are used to provide subjective descriptions of sensory information and to generate a stream of thought. The purpose of such description is to help a user gain a deeper understanding of the role they are playing in the story world.

References

- Bates, J. 1992. Virtual Reality, Art, and Entertainment. *Presence: The Journal of Teleoperators and Virtual Environments* 1(1): pages 133-138.
- Bates, J.; Loyall, A. B.; Reilly, W. S. 1992. Integrating Reactivity, Goals and Emotion in a Broad Agent, Technical Report, CMU-CS-92-142, Department of Computer Science, Carnegie Mellon Univ.
- Galyean, T. A. 1995. Narrative Guidance of Interactivity. Ph.D. Diss, Media Lab, MIT.
- Hayes-Roth, B., and van Gent, R. 1996. Story Making with Improvisational Puppets and Actors, Technical Report, KSL-96-05, Stanford Knowledge Systems Laboratory, Stanford Univ.
- Howe, N., and Strauss B. 1993. *13th Gen: Abort, Retry, Ignore, Fail?*. New York, NY: Vintage Books.
- Kantrowitz, M. and Bates, J. 1992. Integrated Natural

Language Generation Systems, Technical Report, CMU-CS-92-107, Department of Computer Science, Carnegie Mellon Univ.

Kelso, M.T.; Weyhrauch, P.; and Bates, J. 1992. Dramatic Presence. *Presence: The Journal of Teleoperators and Virtual Environments* 2(1): pages.

Loh, K. 1993. Plot summary for Remains of the Day, Internet Movie Database, <http://us.imdb.com/>.

Mueller, E. 1990. *Daydreaming in Humans and Machines*. Norwood, NJ: Ablex.

Neal Reilly, W. S. 1996. Believable Social and Emotional Agents. Ph.D. diss., School of Computer Science, Carnegie Mellon Univ.

Smith, S., and Bates, J. 1989. Towards a Theory of Narrative for Interactive Fiction, Technical Report, CMU-CS-89-121, Dept. of Computer Science, Carnegie Mellon Univ.