

# Toward Characters Who Observe, Tell, Misremember, and Lie

James Owen Ryan, Adam Summerville, Michael Mateas, and Noah Wardrip-Fruin

Expressive Intelligence Studio  
Center for Games and Playable Media  
University of California, Santa Cruz  
{jor, asummerv, michaelm, nwf}@soe.ucsc.edu

## Abstract

Knowledge and its attendant phenomena are central to human storytelling and to the human experience more generally, but we find very few games that revolve around these concerns. This works to preclude a whole class of narrative experiences in games, and it also damages character believability. In this paper, we present an AI framework that supports gameplay with non-player characters who observe and form knowledge about the world, propagate knowledge to other characters, misremember and forget knowledge, and lie. We outline this framework through the lens of a gameplay experience that is intended to showcase it, called *Talk of the Town*, which we are currently developing. From a review of earlier projects, we find that our system has a novel combination of features found only independently across other systems, and that it is among the first to support character memory fallibility.

## Introduction

Nearly all stories rely on characters forming and propagating beliefs about the world, and indeed character belief is a first-class notion in many narratological formalisms, right alongside goals and plans (Elson 2012). But while character goals and plans have often been richly modeled in games, character beliefs seldom have. Beyond formation and propagation, many beloved stories hinge more specifically on knowledge phenomena like false beliefs, memory fallibility, and lies. In *Romeo and Juliet*, the star-crossed lovers take their own lives as a consequence of Romeo’s false belief that Juliet had died; in *The Count of Monte Cristo*, Dantés exacts vengeance on three men who have forgotten his appearance due to the passing of time; and the titular character in *Little Red Riding Hood* fatally believes the wolf’s lie that he is her grandmother. But while knowledge propagation, memory fallibility, and lies are central to human storytelling and to the human experience more generally, we find very few games that revolve around these concerns. This works to preclude a whole class of narrative experiences in games, and it also damages character believability.

When characters do not have knowledge that reactively updates in a way that is consistent with their fields of percep-

tion, we find problematic phenomena like omniscient characters, or characters whose knowledge of the world remains constant even as they witness remarkable in-game events. Because games have not featured actual knowledge propagation, we often encounter the troublesome case of characters in one area of the game world appearing cognizant of the player’s behavior in another area before word of it could have ever spread to them. Lacking support for memory fallibility and lies, we find further unbelievable character behaviors such as perfect recall or the awkward divulging of information in a way that is obviously harmful to the teller.

In this paper, we move toward rectifying this situation by presenting an AI framework that supports gameplay with non-player characters (NPCs) who observe and form knowledge about the world, propagate knowledge to other characters, misremember and forget knowledge, and lie. Subscribing to the stance articulated elsewhere that AI systems intended to support new types of interactive experiences should be appraised through actual, implemented experiences built atop them (Mateas 2001), we have developed this framework in the context of a specific gameplay experience. As such, we present our framework through the lens of this game, *Talk of the Town*, which we are currently developing. Finally, from a fairly comprehensive review of earlier projects, we find that our system has a novel combination of features found only independently across other systems, and that it is among the first to support memory fallibility.

## Design Goals

In developing our framework, we constructed the following design goals:

- Characters form and propagate knowledge about the game world (especially other characters, including player characters) as they go about in it. We wanted to support gameplay with characters who have internal worlds and who are deeply reactive to in-game events.
- Character knowledge may be inaccurate, but for reasons that may in and of themselves be interesting, for instance, lying, misremembering, or unintentionally fabricating false information. We were influenced in this regard by witness false memory.
- The flow of information can be tracked and recalled by the game system. We wanted to support things like visu-

alizing, after gameplay, how an important piece of information originated and propagated across the game world.

Having these goals, we proceeded to design a gameplay experience that would be structured around (and would foreground) the core concerns that they evoke. In the next section, we describe our game, *Talk of the Town*, and thereafter the AI framework that underpins it.

### *Talk of the Town*

*Talk of the Town* is an asymmetric multiplayer dwarflike<sup>1</sup> that features character knowledge propagation as a core mechanic. In this section, we describe its story, simulation, and our gameplay design; the simulation is completed, but the gameplay experience has not yet been implemented.

#### Story

The story that frames gameplay surrounds the death of a very important person who, seventy years prior, founded the town in which gameplay takes place. Since that time, the founder has accumulated considerable wealth and produced several descendants who now constitute an aristocracy in the town. Many of these family members had been anticipating the founder's death for the inevitably large inheritances that would thereby be disbursed, but in his or her<sup>2</sup> last moments the founder apparently signed a document willing everything to a secret lover whose existence had not been known to the family. In one week, the town will gather at a theater to remember the deceased and to hear the will be read, but the family plans to ascertain the identity of the lover and apprehend this person before the document can ever be delivered to the presiding attorney. Meanwhile, the town is abuzz with rumors about the mysterious lover, whom a handful of witnesses briefly observed on the night of the founder's death.

#### Simulation

Prior to gameplay, the town is simulated from its beginnings up through the founder's death seventy years later. Similarly to *Dwarf Fortress* and its world-generation procedure that precedes gameplay (Adams and Adams 2006), this causes a number of structures that are crucial to gameplay to emerge bottom-up from the simulation itself. In *Talk of the Town*, these are the town's physical layout (namely the locations of its businesses and homes), its residents' daily routines, and, most importantly, the city's social and family networks that knowledge propagates over. As we explain more deeply in the next section, NPCs in *Talk of the Town* build up knowledge about their fellow residents (as well as landmarks in the town) through firsthand observation and also by hearing things from other people. This happens online during the simulation, in which characters act out daily routines across day and night timesteps by either going to work, going on errands, visiting friends and family, or staying home;<sup>3</sup> additionally, characters may, for instance, start a business, hire an

<sup>1</sup>A game in the mold of *Dwarf Fortress* (Adams and Adams 2006).

<sup>2</sup>The founder's gender is determined at runtime.

<sup>3</sup>On a given timestep, all characters will be at a specific home or business and will stay there for the duration of the timestep.

employee, build a house, marry another character, give birth to a new character, and so forth. Characters decide what to do by utility-based action selection (Maes 1989). When in the same place, characters will form knowledge about each other through direct observation and might also interact, depending on their relationship and personalities. During interaction, characters may exchange information about the world and, from a simple affinity system, continued interaction may breed contempt, friendliness, or romantic feelings (these work unidirectionally and may be asymmetric). The combinatorics of these simple character behaviors over several in-game decades is enough to generate rich city topologies and social and family networks by the time that gameplay takes place, at which point approximately 200 NPCs will live in the city.

#### Gameplay

Unlike its simulation, *Talk of the Town*'s gameplay has not yet been implemented, but we will describe its design. The game is multiplayer and asymmetric: one player controls the lover character and the other player controls a member of the founder's family. The lover's goal is to go undetected until the will ceremony, while the family member works to ascertain the lover's appearance before that time. (A given character's appearance is the composite of 24 facial attributes, which are inherited from the character's parents.) Because the family character is established in the town, the player controlling him or her will have the town's entire knowledge network at her disposal. As such, her job becomes *managing* this network so that information about the lover's appearance flows toward her; the lover player's task then is to *pollute* this knowledge network by, for instance, changing the character's appearance and spreading lies. Gameplay culminates in a scene showing the town's citizens filing into the theater for the will ceremony, during which time the family player must select the person who best matches her conception of the lover—if she selects correctly, she wins; otherwise, the lover player wins.

### Our AI Framework

Characters in *Talk of the Town* build up knowledge about the world as they go about their simulated daily routines. In this section, we outline our knowledge-representation formalism and the mechanisms by which knowledge may originate, propagate, deteriorate, and terminate according to the procedures of our AI framework, which is written in Python.

#### Ontological Structure

A character's composite knowledge of the world is structured as an *ontology* of interlinked mental models that each pertain to a single person or place. The interlinking occurs when a character's knowledge as to some attribute of a character or place resolves to some other character or place for whom or which they have another mental model. For instance, a character may know that some person works at some business in town, and so her belief about that person's workplace would itself link to her mental model of that business (as we show below). We use this ontological structure for elegance and convenience, since it allows characters to

reason about entities in terms of knowledge they may already have about related entities (rather than by instantiating redundant or potentially inconsistent knowledge).

## Mental Models

As we have alluded to, characters form *mental models* about town residents and landmarks. Following *PsychSim* and *Thespian*, two earlier systems in which characters also richly perceive one another (Marsella, Pynadath, and Read 2004; Si, Marsella, and Pynadath 2005), we use this term specifically to indicate that the structure of a character's knowledge of some entity closely matches the structure of how that entity is represented in the simulation itself. Character mental models pertain to specific individual entities and are composed of *belief facets* that each pertain to an individual attribute of an entity. The facet types that we have implemented so far for *Talk of the Town* match the domain and central concerns of that game and are as follows:<sup>4</sup>

- For mental models of characters:
  - **Name.** First name, middle name, last name.
  - **Appearance.** Each of the 24 facial attributes that we model (*e.g.*, hair color).
  - **Occupation.** Company (links to mental model of that place), job title, shift (day or night).
  - **Home.** Home (either an apartment unit or house; links to mental model of that place).
  - **Whereabouts.** Where a person was on a given day or night (links to mental model of that place).
- For mental models of businesses/homes:
  - **Owner.** Owner of the business/home (links to mental model of that character).
  - **Employees/Residents.** List of its employees/residents (each links to mental model of a character).
  - **Apartment.** Whether it is an apartment unit (for homes only).
  - **Block.** *E.g.*, '900 block of Lake Street'.
  - **Address.** *E.g.*, '147 Hennepin Avenue'.

We plan to extend these to also include other features of characters and landmarks that are already modeled by the simulation (such as character age, personality, family members, friends, enemies, etc.).

Each facet is structured as a collection of data about the belief. In addition to its *owner* (the character who has constructed the mental model), *subject* (the character to whom it pertains), and facet type, this data includes:

- **Value.** A representation of the belief itself, *e.g.*, the string 'brown' for a belief facet pertaining to hair color.
- **Mental Model.** If the value of this facet resolves to an entity for whom the owner of this facet has formed a mental model, this will point to that mental model.
- **Predecessor.** The belief facet that the owner previously held, if any. This allows the system to track supplanted or forgotten character knowledge.

<sup>4</sup>Note, however, that our knowledge-representation formalism is agnostic to the type of knowledge that it is used to represent.

- **Evidence.** A list of the pieces of evidence by which the owner of this facet formed and continues to substantiate it; evidence may accumulate as the simulation proceeds. In the next section, we outline our evidence typology.
- **Strength.** The strength of this particular belief. This is the sum of the strength of all pieces of evidence supporting this belief, whose determination we explain in the next section.
- **Accuracy.** Whether or not the belief is accurate (with regard to the *current* true state of the world).

## Evidence

All knowledge gets formed in response to evidence, and may also propagate, deteriorate, or terminate in a way that can be described using pieces of evidence. We will illustrate these details by explaining our evidence typology, which comprises nine *types* across four categories.

- How knowledge originates:
  - **Reflection.** A reflection occurs when a character perceives something about herself, which happens at every timestep.
  - **Observation.** When a character directly observes a person or place, she may form knowledge about attributes of that entity. Whether she forms knowledge about a particular attribute depends on the salience of the entity and the attribute type, as we explain below.
  - **Transference.** If one entity reminds a character of another entity (determined by feature overlap between her respective mental models of them), she may unconsciously attribute beliefs she already held about one to the other.
  - **Confabulation.** By confabulation, a character *unintentionally* concocts new knowledge about some entity. The particular belief-facet value that gets confabulated is determined probabilistically according to the distribution of that feature type in the town.
  - **Lie.** A lie occurs when a character *intentionally* conveys information to another character that she herself does not believe. We call this a type of origination (and not propagation) because the knowledge in question is *invented* by virtue of the lie—*i.e.*, no existing knowledge is propagated by the lie.
- How knowledge propagates:
  - **Statement.** A statement occurs when a character conveys information to another character that she herself believes. Whether characters will exchange a particular piece of information depends on the salience of its subject and type, as we discuss below.
  - **Eavesdropping.** Nearby characters may overhear statements and lies; this happens at a set probability.
- How knowledge deteriorates:
  - **Mutation.** As an operationalization of memory fallibility, knowledge may mutate over time; this is affected by a character's memory attribute (which is modeled as a floating-point value that gets inherited from a

parent) and the facet type (e.g., a *whereabouts* belief will be more likely to mutate than a *first name* belief). The particular mutation that occurs is determined by a schema we authored that specifies state-change probabilities given a facet value. For instance, given the value ‘brown’ for the facet type *hair color*, the system will consult a (hand-authored) probability distribution for which, e.g., ‘black’ and ‘red’ would be more probable values to mutate to than ‘white’ or ‘gray’.

- How knowledge terminates:
  - **Forgetting.** To further incorporate memory fallibility, knowledge may be forgotten due to time passing; this is likewise affected by a character’s memory attribute and the salience of the facet subject and type.

Characters are not consciously aware of transferences, confabulations, or mutations, and recipients (and eavesdroppers) of lies treat them as statements. That is, the recipient will reason about a lie as if it were a statement (and so the strength of a lie, as a piece of evidence, is equal to that of a statement), but the system will still track that it was in fact a lie, to allow for the presentation of true knowledge trajectories after gameplay. Additionally, each piece of evidence has metadata of the following types:

- **Source.** With a statement, lie, or eavesdropping, this specifies the character who delivered the information. This crucially allows the system to trace the history and trajectory of any piece of information, which was one of the design goals that we gave above.
- **Location.** Where the piece of evidence originated (e.g., where an observation or statement took place).
- **Time.** The timestep a piece of evidence originated.
- **Strength.** The strength of a piece of evidence is a floating-point value that is determined by its type (e.g., a mutation is weaker than an observation) and decays as time passes. In the case of statements, lies, and eavesdroppings, the strength of a piece of evidence is also affected by the affinity its owner has for its source and the strength of that source’s own belief at the time of propagation.

## Salience Computation

When a character observes some entity in the simulation, a procedure is enacted that determines, for each perceptible attribute of the observed entity,<sup>5</sup> the probability that the character will remember what she saw; this procedure crucially depends on the *salience* of the entity and attribute being observed. At this time, salience computation considers the relationship of an observed character (subject) to the observer (e.g., a co-worker is more salient than a stranger), the extent of the observer’s friendship with the subject, the strength of the observer’s romantic feelings toward the subject, and finally the subject’s job level (characters with more prestigious job positions are treated as more salient). For landmarks, salience computation currently only considers

<sup>5</sup>E.g., physical attributes, a character’s workplace and shift if they are observed at work, etc.

whether the observing character lives or works at the observed place. Additionally, our salience-computation procedures consult a hand-authored knowledgebase specifying the salience of each attribute type. This captures, for instance, that features of a person’s hair and eyes will be more salient than those of her nose and chin (Reynolds and Pezdek 1992; Ruiz-Soler and Beltran 2012). Salience computation is also used to determine the probability that a character will misremember or altogether forget (on some later timestep) knowledge pertaining to some subject and attribute.

## Knowledge Propagation

The salience of the subject and attribute type of a piece of information also affects whether a character will pass it on (via a statement). Currently, what subjects of conversation come up in an interaction between two characters is determined by computing the salience of all entities that either of the characters know about. The  $n$  highest-scoring entities are then brought up in conversation, with  $n$  being determined by the strength of the characters’ relationship and also their respective *extroversion* personality components. For each subject of conversation, the characters will exchange information about individual attributes of that subject (corresponding to the individual belief facets of the characters’ mental models of that subject) according to the salience of each attribute type. Because a character may bring up subjects that her interlocutor does not (yet) know about, our propagation mechanism allows characters to learn about other people and landmarks that they have never encountered themselves. It is even possible for a character to learn about another character who died before she was born; this often occurs when parents tell their children about deceased relatives.

## Lies

As a subject of conversation gets brought up, a character may convey false information—more precisely, information that she herself does not believe—to her interlocutor. Currently, this happens probabilistically according to a character’s affinity toward the interlocutor, and the misinformation is randomly chosen. Later, we discuss plans to extend this aspect of the system.

## Belief Revision

Currently, characters will always adopt a new belief upon encountering a first piece of evidence supporting it, assuming they have no current belief that it would replace. As a character accumulates further evidence supporting her belief, its strength will increase commensurately to the strength of the new evidence. Additionally, whenever a character delivers a statement, the strength of her own belief (that she is imparting with the statement) will slightly increase. That is, the more a person retells some belief, the stronger that belief becomes for her, which is realistic (Wilson, Gambrell, and Pfeiffer 1985). If that character, however, encounters new evidence that contradicts her currently held view, she will consider the strength of the new evidence relative to the strength of her current belief. If the new evidence is stronger, she will adopt the new belief that it supports; if it is weaker,

she will *not* adopt a new belief, but will still keep track of the other *candidate belief* and the evidence for it that she had encountered. If she continues to encounter evidence supporting the candidate belief, she will update its strength accordingly and if at any time that strength exceeds the strength of her current belief, she will adopt the candidate belief and relegate the previously held belief to candidate status.

As an example, consider two characters, Jack and Jill. Jill has just gotten a haircut, but Jack still believes that Jill has long hair because he had seen this for himself on multiple occasions and heard the same from a good friend just yesterday. During a conversation, another character that Jack does not know very well mentions to him that Jill has short hair. Jack does not immediately change his belief about Jill's hair, because the strength of this new evidence does not outweigh the strength of his current belief (since he had heard only yesterday that she has long hair and saw this himself on multiple occasions). Later, Jack's wife attests to Jack that Jill's hair is in fact short. Jack now adopts the new belief, because the strength of the evidence constituted by both of these statements now outweighs the strength of the evidence supporting his prior belief. The prior belief was supported by more pieces of evidence, but this evidence was so outdated relative to the new information that its strength had decayed to the point where the new evidence was now stronger.

### Satisfaction of Design Goals

We invite the reader to now revisit the design goals that we enumerated earlier in this paper. The system as it is currently implemented is already capable of satisfying each of these goals, though we are now faced with considerable authoring challenges in getting character knowledge to the surface in *Talk of the Town* gameplay.

### Implementation Status

Our major outstanding work is to implement the actual gameplay experience of *Talk of the Town*. The underlying simulation, as we have described it, is already implemented, though we envision several improvements. First, there is a significant amount of work that remains to be done involving modeling how characters may reason about the solicitation and divulging of information. At this time, characters do not explicitly solicit or divulge information—there is only a notion of subjects of conversation getting talked about according to how salient they are to both conversational partners. As part of fleshing out this aspect of the system, a particularly rich extension will involve implementing interesting character reasoning about when to lie and what to lie about. We already model several components of lying behavior that are described in social-science research, including personality, quality of social and family relationships, and social standing (Kashy and DePaulo 1996)—the work will be to construct interesting character behavior that depends on these components. Currently, lies are only stored in the knowledge of characters who receive them, but we plan to have characters who tell them also keep track of them so that they can reason about past lies when constructing subsequent ones. While characters currently only lie about other

characters, we plan to also implement *self-centered* lying (DePaulo 2004), *e.g.*, characters lying about their job titles or relationships with other characters. Finally, we envision characters who discover they have been lied to revising their affinities toward the liars, or even confronting them.

### Prior Work and Discussion

While story generators and expressive multiagent systems have typically featured characters who operate over perfect knowledge of the world (*e.g.*, Aylett, Dias, and Paiva 2006; Fendt and Young 2011), a handful of such systems have incorporated models of character belief. In TALE-SPIN, characters may be initialized to have some knowledge about the storyworld and may perceive which characters and objects are nearby (Meehan 1976). Agents in applications of the *Oz Project* similarly perceive their surroundings, but at a higher fidelity (Bates, Loyall, and Reilly 1994). This is likewise seen in ten Brinke, Linssen, and Theune's (2014) extension to the *Virtual Storyteller* system (Swartjes 2010), in which updates to character beliefs are fed to a reactive narrative-planning system. Teutenberg and Porteous (2015) have also explored narrative planning with perceptive characters. Gervás (2013) uses chess gameplay data to generate stories focalized to individual pieces with limited *fields of perception*. In the emergent-narrative system of Carvalho et al. (2012), characters may operate over false beliefs. In TALE-SPIN, *PsychSim* (Marsella, Pynadath, and Read 2004), *Thespian* (Si, Marsella, and Pynadath 2005), *Othello* (Chang and Soo 2009), and Reis's (2012) extension to the *FAtiMA* agent architecture (Aylett, Dias, and Paiva 2006), characters not only perceive the world, but form mental models of other characters and *their* beliefs—*i.e.*, these systems operationalize *theory of mind* (Frith and Frith 2005). Earlier, we proposed this as an extension to our own system, which also features character mental models. Further, a handful of multiagent systems have explored *deception in agents*. TALE-SPIN characters may lie to one another (Meehan 1976, 183-84), though rather arbitrarily, as in our current system implementation. GOLEM implements a *blocks world* variant in which agents deceive others to achieve goals (Castelfranchi, Falcone, and De Rosis 1998), while *Mouth of Truth* uses a probabilistic representation of character belief to fuel agent deception in a variant of Turing's *imitation game* (De Rosis et al. 2003). In Christian (2004), a *deception planner* injects inaccurate world state into the beliefs of a target agent so that she may unwittingly carry out actions that fulfill ulterior goals of a deceiving agent. Lastly, agents in Reis's (2012) extension to *FAtiMA* employ multiple levels of theory of mind to deceive one another in the party game *Werewolf*. While all of the above systems showcase characters who perceive—and in some cases, deceive—other characters, none appear to support the following key components of our system: knowledge propagation and memory fallibility.

Outside of *Dwarf Fortress* (Adams and Adams 2006), which we discuss next, commercial and even research games have been surprisingly spare in their modeling of character knowledge, as we recount here. Combat-oriented games often employ a rudimentary notion of character knowl-

edge reasoning in the form of NPC sensors for (realistically imperfect) detection of, e.g., enemy positions. *Reputation systems* or *faction systems* that alter NPC behavior toward the player character according to actions she has taken are a form of character knowledge representation, but one in which NPCs seem to globally inherit perfect knowledge of player behavior. *Black and White*'s "creatures" were special NPCs that could acquire knowledge (by reinforcement learning) about how to do certain tasks (Lionhead Studios 2014). Interestingly for a mainstream title, the recent *Middle-earth: Shadow of Mordor* features NPC enemies who may reference earlier combat (Monolith Productions 2014). More commonly, games may utilize simple *flags* that mark what plot knowledge the player currently has, so that only valid dialogue (sub)trees get deployed (Wardrip-Fruin 2009). In an extension to *Comme il Faut* (McCoy et al. 2014)—the AI system underpinning *Prom Week*, which features omniscient characters (McCoy et al. 2013)—Sullivan et al. (2012) likewise incorporate a flag-like representation of player plot knowledge, but also a more general notion of character world knowledge (which specifies things like characters' knowledge of other characters' traits). Most interestingly, in this system the player may manipulate NPC knowledge by asserting facts about the world, but this information does not propagate because the NPCs do not then interact with one another in this same way. Finally, in *Versu*, characters may have beliefs that pertain to specific story concerns and, through dialogue interaction, may even persuade other characters to adopt such beliefs (Evans and Short 2014). But because *Versu*'s AI framework has only been applied to storyworlds with a small number of characters, it is probably not accurate to say that it features knowledge propagation; moreover, it does not appear to operationalize memory fallibility or character lies.

It seems that *Dwarf Fortress* (Adams and Adams 2006) best represents the state of the art (certainly in the domain of digital games) in modeling character knowledge and its attendant phenomena. The information that follows was gathered from an email correspondence with its creator, Tarn Adams (May 28, 2015). The primary character-knowledge machinery in *Dwarf Fortress* is its *rumors system*. Rumors typically originate when a character witnesses an important event, such as a crime, and are instantiated at certain local and regional levels. Interestingly, characters in the purview of a rumor might then propagate it to new areas they visit. Examples of this include diplomats informing their home civilizations about traps they have encountered, caravans (who visit player fortresses) bringing rumors from their parent civilizations, and artistic and scientific knowledge moving across civilizations. Like a few other systems noted above, *Dwarf Fortress* also features characters who autonomously lie. When a character commits a crime, she may falsely implicate someone else in a witness report to a sheriff, to protect herself or even to frame an enemy. These witness reports, however, are only seen by the player; characters don't give false witness reports to each other. They may, however, lie about their opinions, for instance, out of fear of repercussions from criticizing a leader. Finally, *Dwarf Fortress* does not currently model issues of memory

fallibility—Adams is wary that such phenomena would appear to arise from bugs if not artfully expressed to the player.

Lastly, we note that CYRUS, an AI system that operationalizes reconstructive memory, exhibits memory fallibility, though it is not a multiagent system (Kolodner 1983).

Our approach exhibits several features that are found only independently across the various systems that we have outlined. As in systems like *PsychSim* and *Thespian*, characters in ours actively form mental models of one another. Following GOLEM, *Mouth of Truth*, *Dwarf Fortress*, and a few other projects, our framework supports character lying, though we currently do not model this aspect nearly as richly. Along with *Dwarf Fortress*, our system is among the first to support character knowledge propagation. Lastly, to our knowledge, ours is the first multiagent system to feature fallible character memory. Because all these phenomena may interact and combine in several ways, our system yields character-knowledge topologies whose richness appears unmatched by those of earlier systems.

## Conclusion and Future Work

Knowledge and its attendant phenomena are central to human storytelling and to the human experience more generally, but we find very few games that revolve around these concerns. In this paper, we presented an AI framework that supports gameplay with NPCs who observe and form knowledge about the world, propagate knowledge to other characters, misremember and forget knowledge, and lie. From a review of earlier projects, we find that our system has a novel combination of features found only independently across other systems, and that it is among the first to support memory fallibility.

While we have developed this framework in the context of a particular gameplay experience, *Talk of the Town*, we anticipate compelling applications of it beyond the purview of that experience. Specifically, we envision implementations made possible by the fact that our representation formalism is agnostic to the type of knowledge that it is used to represent. For instance, one could have characters build up *metaknowledge* simply by representing their knowledge about their own knowledge (e.g., the source, location, and time of a statement a character received) as first-class belief facets that originate, propagate, deteriorate, and terminate just as we have described in this paper. More potently, we could operationalize theory of mind by likewise instantiating as facets beliefs about *other* character's beliefs.<sup>6</sup> While we have noted such operationalization in three narrative planning systems (Marsella, Pynadath, and Read 2004; Si, Marsella, and Pynadath 2005; Chang and Soo 2009), we believe it could also support novel gameplay experiences in which, for instance, NPCs actively reason about the player's knowledge (by reasoning about her character's knowledge). Lastly, beyond these extensions, we believe our knowledge-representation formalism could accommodate beliefs that are more subjective in nature (or abstract in what they represent) than facts about the world. For instance, if belief

<sup>6</sup>Or even a character's beliefs about another character's beliefs about her beliefs, as Reis (2012) explored.

facets were to represent sociocultural concerns, our framework could be used to model sociocultural transmission and evolution (Cavalli-Sforza and Feldman 1981).

## Acknowledgments

We thank Tarn Adams for graciously providing a detailed account of knowledge phenomena in *Dwarf Fortress*, and Jacob Garbe, Ian Horswill, and Jonathan Lessard for their kind words and encouragement. We also thank our anonymous reviewers for their helpful feedback.

## References

- Adams, T., and Adams, Z. 2006. *Slaves to Armok: God of Blood Chapter II: Dwarf Fortress*. Bay 12 Games.
- Aylett, R.; Dias, J.; and Paiva, A. 2006. An affectively driven planner for synthetic characters. In *ICAPS*, 2–10.
- Bates, J.; Loyall, A. B.; and Reilly, W. S. 1994. *An architecture for action, emotion, and social behavior*. Springer.
- Carvalho, D. B.; Clua, E. G.; Passos, E. B.; and Pozzer, C. T. 2012. A perception simulation architecture for plot generation of emergent storytelling. *Computer Games, Multimedia, and Allied Technology* 6.
- Castelfranchi, C.; Falcone, R.; and De Rosis, F. 1998. Deceiving in golem: how to strategically pilfer help. 91–109.
- Cavalli-Sforza, L. L., and Feldman, M. W. 1981. *Cultural transmission and evolution: a quantitative approach*. Number 16. Princeton University Press.
- Chang, H.-M., and Soo, V.-W. 2009. Planning-based narrative generation in simulated game universes. *IEEE Transactions on Computational Intelligence and AI in Games*.
- Christian, D. 2004. Strategic deception in agents. Master's thesis, North Carolina State University.
- De Rosis, F.; Carofiglio, V.; Grassano, G.; and Castelfranchi, C. 2003. Can computers deliberately deceive? a simulation tool and its application to turing's imitation game. *Computational Intelligence* 19(3):235–263.
- DePaulo, B. M. 2004. The many faces of lies. *The social psychology of good and evil*, ed. AG Miller 303–26.
- Elson, D. K. 2012. *Modeling narrative discourse*. Ph.D. Dissertation, Columbia University.
- Evans, R., and Short, E. 2014. Versu—a simulationist storytelling system. *IEEE Transactions on Computational Intelligence and AI in Games* 6(2):113–130.
- Fendt, M. W., and Young, R. M. 2011. The case for intention revision in stories and its incorporation into iris, a story-based planning system. In *Intelligent Narrative Technologies*.
- Frith, C., and Frith, U. 2005. Theory of mind. *Current Biology* 15(17):R644–R645.
- Gervás, P. 2013. Stories from games: Content and focalization selection in narrative composition. In *Actas del Primer Simposio Español de Entretenimiento Digital*.
- Kashy, D. A., and DePaulo, B. M. 1996. Who lies? *Journal of Personality and Social Psychology* 70(5):1037.
- Kolodner, J. L. 1983. Reconstructive memory: A computer model. *Cognitive science* 7(4):281–328.
- Lionhead Studios. 2014. *Black and White*. Electronic Arts.
- Maes, P. 1989. How to do the right thing. *Connection Science* 1(3):291–323.
- Marsella, S. C.; Pynadath, D. V.; and Read, S. J. 2004. Psychsim: Agent-based modeling of social interactions and influence. In *Proceedings of the International Conference on Cognitive Modeling*, volume 36, 243–248.
- Mateas, M. 2001. Expressive ai: A hybrid art and science practice. *Leonardo*.
- McCoy, J.; Treanor, M.; Samuel, B.; Reed, A. A.; Mateas, M.; and Wardrip-Fruin, N. 2013. Prom week: Designing past the game/story dilemma. In *Proc. Foundations of Digital Games*, 94–101.
- McCoy, J.; Treanor, M.; Samuel, B.; Reed, A.; Mateas, M.; and Wardrip-Fruin, N. 2014. Social story worlds with comme il faut. *IEEE Transactions on Computational Intelligence and AI in Games PP (99)* 1–1.
- Meehan, J. R. 1976. *The metanovel: writing stories by computer*. Ph.D. Dissertation, Yale University.
- Monolith Productions. 2014. *Middle-earth: Shadow of Mordor*. Warner Bros. Interactive Entertainment.
- Reis, H. D. S. 2012. Lie to me: Lying virtual agents. Master's thesis, Universidade Técnica de Lisboa.
- Reynolds, J. K., and Pezdek, K. 1992. Face recognition memory: The effects of exposure duration and encoding instruction. *Applied Cognitive Psychology* 6(4):279–292.
- Ruiz-Soler, M., and Beltran, F. S. 2012. The relative salience of facial features when differentiating faces based on an interference paradigm. *Journal of Nonverbal Behavior* 36(3):191–203.
- Si, M.; Marsella, S. C.; and Pynadath, D. V. 2005. Thespian: Using multi-agent fitting to craft interactive drama. In *Proc. ICAAMS*, 21–28. ACM.
- Sullivan, A.; Grow, A.; Mateas, M.; and Wardrip-Fruin, N. 2012. The design of mismanor: creating a playable quest-based story game. In *Proceedings of the International Conference on the Foundations of Digital Games*, 180–187.
- Swartjes, I. 2010. *Whose Story Is It Anyway: How Improv Informs Agency and Authorship of Emergent Narrative*. Ph.D. Dissertation, University of Twente.
- ten Brinke, H.; Linssen, J.; and Theune, M. 2014. Hide and sneak: Story generation with characters that perceive and assume. In *Proc. AIIDE*.
- Teutenberg, J., and Porteous, J. 2015. Incorporating global and local knowledge in intentional narrative planning. In *Proc. ICAAMS*, 1539–1546.
- Wardrip-Fruin, N. 2009. *Expressive Processing: Digital fictions, computer games, and software studies*. MIT press.
- Wilson, R. M.; Gambrell, L. B.; and Pfeiffer, W. R. 1985. The effects of retelling upon reading comprehension and recall of text information. *The Journal of Educational Research* 78(4):216–220.