Who Knows What?
Perspective-Enabled Story Understanding

by

Jessica Marie Noss

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHVSETTS INSTITVTE OF TECHNOLOGY

February 2017

Author .....
Department of Electrical Engineering and Computer Science
February 3, 2017

Certified by .....
Patrick H. Winston
Ford Professor of Artificial Intelligence and Computer Science
Thesis Supervisor

Accepted by .....
Christopher J. Terman
Chairman, Masters of Engineering Thesis Committee

Who Knows What?

Perspective-Enabled Story Understanding

by

Jessica Marie Noss

Submitted to the Department of Electrical Engineering and Computer Science
on February 3, 2017, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Stories are essential to human communication, thus understanding stories is critical to modeling human intelligence. In order to understand a story from multiple characters' perspectives, a reader must form mental models representing each character's knowledge and beliefs. Building on the Genesis story-understanding system, I developed the Perspectives Expert, which enables Genesis to form detailed mental models of characters in a story. I started by annotating stories with characters' entrances and exits, treating stories as theatrical plays. I assumed that characters observe events only while onstage, and I used their observations to populate mental models. Then, inspired by human reasoning, I eliminated explicit stage directions and enabled Genesis to keep track of characters' locations and levels of consciousness. Equipped with this mental-modeling ability, Genesis can retell stories from different characters' perspectives, answer reading-comprehension questions to demonstrate understanding, and perform the first steps of conflict reconciliation by detecting and explaining opposing viewpoints.

Thesis Supervisor: Patrick H. Winston
Title: Ford Professor of Artificial Intelligence and Computer Science

Acknowledgments

Without the support of many incredible people, I could not be where, what, and who I am today. In particular:

Patrick Winston, for showing me how to write, speak, and teach; for the constant inspiration, support, and empowerment; for the invaluable stories and life lessons; and for always believing in me.

All of my friends and colleagues in the **Genesis group**—especially Matthew, Sila, Robert, Adam, Igor, Dylan, Ryan, Josh, Brian, Nick, Ernie, Ada, Bryan, Caroline, and Jake—for providing encouragement; for showing me my own progress when I couldn't see it myself; and for all of the thought-provoking discussions on the mechanisms behind human story-understanding.

The **6.034 TAs** from Fall 2014, 2015, and 2016—especially Dylan, Ryan, Neil, Sarah, Veronica, Brittney, Jake, and Rebecca—for supporting me through three exciting semesters full of good decisions (rewriting the labs...), bad decisions (...during the semester), and over a dozen of the best parties I've attended at MIT. (Grading parties, of course!)

The **6.xxx TAs and students** from Spring 2014, 2015, and 2016—especially Theodora, Rebecca, and Nicole—for helping me learn to read, write, and communicate more effectively.

Techiya and **MITOC**, for all the fun times, and for giving me the opportunity to learn how to lead.

Honorable Menschen, for giving me a good reason to leave campus once a week, and for keeping me connected to the "real world" during my final semester of MEng.

The **MIT Musical Theatre Guild** and the **cast of Hack Punt Tool** 2017 revival, for supporting me and filling in for me after my silly decision to act, dance, and sing in a show that opened on the day my thesis was due, and for providing some much appreciated comic relief during my last couple weeks of thesis-writing.

Nate Zuk, for countless hours of companionship, friendship, and music, and always being there (including to help with last-minute proofreading).

Dylan Holmes, for believing in me and encouraging me; for countless hours of unique conversations; and for being a wonderful TA, mentor, role model, discussion companion, and friend.

Rebecca Kekelishvili, for being an incredible copy editor!

Joel Moses, my academic advisor, for allowing me to chart my own unusual course through MIT for more semesters than either of us anticipated.

Chris Terman, for supporting me when I transferred into Course 6 as a junior, and for enabling me to stay at MIT for a 5th term of MEng and continue to be a 6.034 TA.

My family—my parents, Joy and Roger, and my brother, Oliver—for your unconditional love and support, for letting me chart my own course, and for always being there on the other end of the phone.

All of the teachers, friends, and mentors who helped me get to where I am today: The list is far too long to enumerate here.

Ariel Sommer, for supporting me through eleven crazy semesters at MIT, for adventuring around the world with me by plane, train, bus, car, sailboat, bicycle, and foot... and for being the best partner a person could ask for.

And finally—once again—**Patrick Winston**, for teaching me that I could do it, and only I could do it, but I couldn't do it alone. I feel incredibly lucky and proud to have been one of your people.

Contents

1 Introduction **11**

1.1 My Vision 11

1.2 Steps toward perspective-enabled story understanding 13

1.3 What Genesis can do now, as of February 2017 14

1.4 What you will see in the rest of this thesis 15

2 The Genesis Story-Understanding System, circa 2014 **19**

3 Forming Mental Models **25**

3.1 Treating stories as simplified theatre makes characters' observations explicit 26

3.2 Inferring observations from context eliminates stage-direction annotations 30

3.3 Locations help the reader keep track of characters 34

3.4 Unconscious characters cannot observe events 39

3.5 Personal Rules: Individuals use different assumptions about the world 45

3.6 Real stories have complex scenes: Partial information and additional complications 48

4 Applications of Mental Models and Perspectives **51**

4.1 Retelling the story from a character's perspective 51

4.2 Answering reading-comprehension questions 53

4.3 Conflict detection and explanation: Steps toward conflict reconciliation 57

5 Contributions **61**

List of Figures

1-1 Demonstration of Genesis's Perspectives Expert 16

2-1 Sample elaboration graph: *Macbeth* 20
2-2 Revenge concept pattern in *Macbeth* 20
2-3 Revenge concept pattern in Russo-Estonia Cyberwar 22
2-4 Alternate concept pattern (teaching a lesson) in Russo-Estonia Cyberwar 23

3-1 Stories as simplified theatre: Full elaboration graph 27
3-2 Stories as simplified theatre: Jean Valjean's mental model 27
3-3 Stories as simplified theatre: Inspector Javert's mental model 28
3-4 Locations: Full elaboration graph 36
3-5 Locations: Jean Valjean's mental model 36
3-6 Locations: Inspector Javert's mental model 37
3-7 Unconsciousness: Full elaboration graph 41
3-8 Unconsciousness: Enjolras's mental model 43
3-9 Unconsciousness: Marius's mental model 43
3-10 Unconsciousness: Jean Valjean's mental model 44
3-11 Unconsciousness: Thenardier's mental model 44
3-12 Personal rules: Full elaboration graph 47
3-13 Personal rules: Inspector Javert's mental model 49
3-14 Personal rules: Jean Valjean's mental model 49

4-1 Genesis interface with text box 52

Chapter 1

Introduction

1.1 My Vision

Imagine a world in which there is no need to work for a living, and no fear of starvation; a world in which we have cured sickness and death; a world in which we can travel to human settlements on distant planets, and have world peace on Earth. I believe that these lofty goals will be far easier to achieve with the help of intelligent machines—capable of performing tasks ranging from menial labor to advanced research—and one promising path to the engineering goal of developing intelligent machines is the scientific goal of understanding human intelligence.

Story understanding is crucial to understanding human intelligence because humans think in terms of stories, as evidenced by the ubiquity of stories in human cultures, ranging from fictional tales such as fables and plays to nonfiction writings such as news articles, historical accounts, and case studies. Understanding how humans create, tell, and think about stories will help us develop a computational understanding of human intelligence, which can then enable engineering applications such as intelligent machines that interact naturally with humans.

Accordingly, for my thesis, I took an important step toward understanding human intelligence by modeling how humans understand stories from multiple perspectives through mental modeling. I built upon the Genesis story-understanding system, an artificial-intelligence platform for understanding and telling stories, developed by

Patrick Winston's Genesis group [Winston, 2011, Winston, 2012a, Winston, 2012b].

When I joined the Genesis group in 2014, the Genesis system already had a wide range of capabilities to represent stories and model story understanding, ranging from question answering to concept identification to culturally specific points of view. These capabilities allowed it to complete a variety of tasks mimicking how humans understand stories. For instance, given a simplified-English synopsis of Shakespeare's *Macbeth*, Genesis could identify the concepts of *revenge* and *Pyrrhic victory* in the story, even though neither term is explicitly mentioned in the story.

However, when I started my project in 2014, Genesis was missing an important piece of the puzzle: It could not yet construct mental models of characters. Genesis could understand characters' motives from pre-defined points of view, but it was unable to construct a model of an individual character's knowledge of other characters and of the world. Instead, its limited concept of a mental model was merely a place to list a character's personality traits, such as *evil* or *greedy*. Genesis needed the ability to form mental models of characters to represent the characters' views of the world, including their knowledge and interpretation of events that occurred in a story, as well as their rules and assumptions about how the world works.

This was a notable hole in the Genesis system because one important way that humans understand stories is by considering events from different characters' perspectives, in order to understand the rationale behind those characters' actions. In real life, humans also form mental models of each other in order to develop expectations about how others will react to a hypothetical situation. In fact, the ability to form mental models of others—to think about others' thoughts, often referred to as *theory of mind*—is essential to human interaction [Saxe and Kanwisher, 2003] and is often considered to be a uniquely human characteristic [Call and Tomasello, 1999], which makes it even more essential that an artificially intelligent system be able to produce similar behavior.

Perspective-taking, or the process of understanding a situation from another perspective has broader engineering applications beyond fictional story understanding, such as providing useful insight into reconciliation processes. In order to reach a

reconciliation, each side must first be able to understand and accept the legitimacy of the other side's perspective, so a mediator must have the ability to understand the situation from both parties' perspectives [Cobb, 2013]. (And if we are to achieve world peace, reconciliation will certainly be an important step.)

Other possible applications of mental modeling include the ability to retell a story from another character's perspective and the ability to answer reading-comprehension questions about individual characters' thought processes.

Mental models are a powerful tool for understanding stories, and ultimately for understanding the world. Thus, in order to understand stories from multiple perspectives, Genesis must be able to form mental models representing each character's knowledge and beliefs.

1.2 Steps toward perspective-enabled story understanding

In carrying out my research, I:

1. **Developed a representation for mental modeling.** Each character should have their own mental model, containing everything they know or believe about their world. In Genesis, a mental model is a framework that can be populated with the subset of events that a particular character observes or infers—essentially the character's personal story.

2. **Populated mental models by annotating stories.** To simplify the problem of inferring what each character observes, I treated stories as theatrical plays by annotating them with explicit scene markers and characters' entrances and exits, which made it easy to determine which events each character observed.

3. **Inferred observations from context.** If a character performs an action in a scene, Genesis can infer that the character is present, thereby eliminating the need to explicitly annotate a story with characters' entrances and exits.

4. **Tracked characters' locations.** Humans form models of locations and the people they expect to find in those locations. Similarly, Genesis should keep track of characters' locations to enable parallel scenes in a story ("Meanwhile, in the kitchen..."), which requires remembering which characters were in each location. Location tracking can also eliminate explicit scene markers by replacing them with physical changes of location.

5. **Incorporated unconsciousness.** If a character falls asleep, dies, or otherwise becomes unconscious, they should not observe any events that transpire, but they should resume observing if they awaken.

6. **Employed personal rules to draw conclusions.** Some characters have different assumptions and rules about how the world works. For example, one might believe that people shout because they are angry, whereas another believes that people shout because they are excited. These differing rules may cause the characters to reach different conclusions.

7. **Applied Genesis's new mental-modeling capability to retell stories, answer questions, and resolve conflicts.** Using the information stored in each character's mental model, Genesis can retell a story from the point of view of a character in the story, answer reading-comprehension questions to demonstrate understanding, and even perform the first steps in conflict reconciliation: identifying and explaining a disagreement between two characters.

1.3 What Genesis can do now, as of February 2017

Through my work, the Genesis story-understanding system is now able to:

- Form a detailed mental model of every character in a story, incorporating each character's personal assumptions as well as a set of global common-sense rules.

- Identify which characters are present at every moment in a story.

- Recognize when characters lose and regain consciousness (fall asleep, die, become unconscious, awaken), and understand that unconscious characters cannot form observations.
- Keep track of characters' locations throughout a story.
- Retell a story from an individual character's point of view.
- Answer reading-comprehension questions such as:
  - Why does Inspector Javert believe that Jean Valjean committed a crime? (Figure 1-1)
  - Does Inspector Javert believe that Jean Valjean is a criminal?
  - Does Marius think that Jean Valjean is honest?
  - Why did Thenardier believe that Jean Valjean was a murderer?
  - Why does Jean Valjean think that Cosette loves Marius?
- Determine whether two characters disagree, and if so, explain the source of their conflict.

1.4 What you will see in the rest of this thesis

In Chapter 2, I provide a brief overview of Genesis's capabilities prior to my work, to explain the system upon which I was building. In Chapter 3, I describe in detail the steps I took to enable Genesis to form mental models. For each added ability, I start by explaining the motivation for the ability, and then I give a short example story and demonstrate Genesis's results. Finally, I explain how the ability works and some of the considerations that went into it, and I compare and contrast Genesis's ability with human behavior. In Chapter 4, I present the three applications of mental modeling that I implemented: Retelling, question answering, and conflict detection and explanation. For each application, I provide the motivation, an example, and

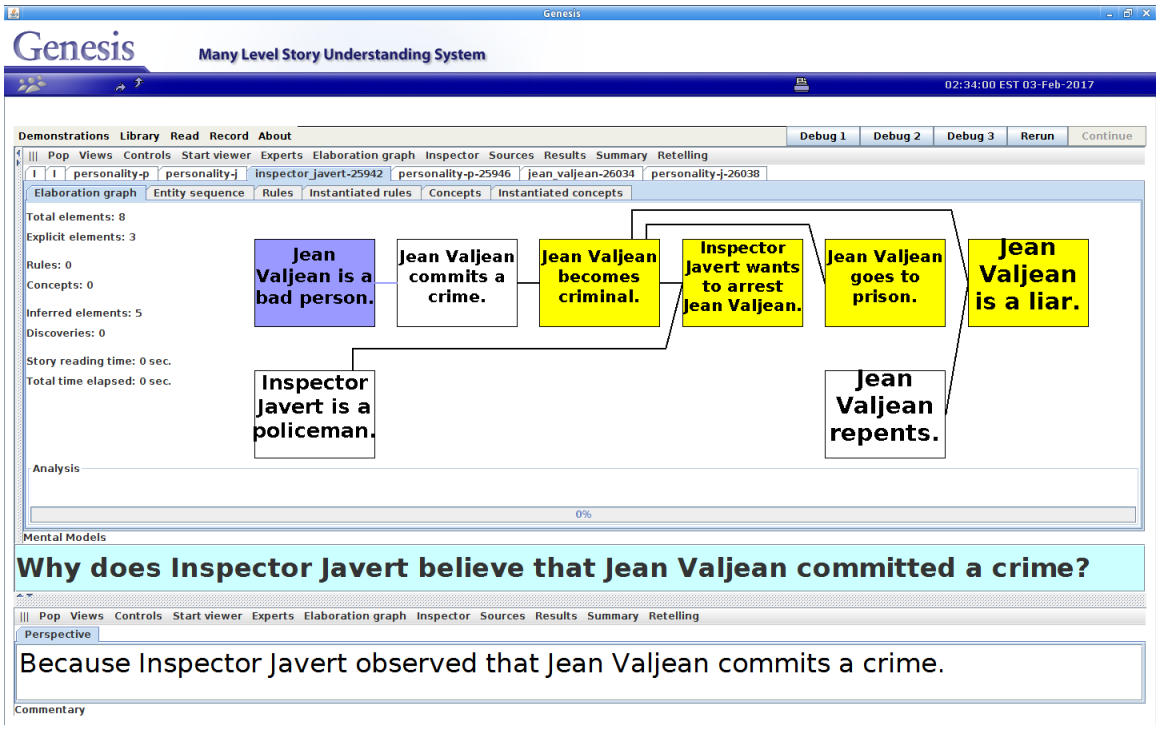

Figure 1-1: Genesis's Perspectives Expert forms mental models of characters' observations and beliefs, and it uses those mental models to answer questions about characters' knowledge.

a description of how the application works. Finally, in Chapter 5, I summarize the main contributions of my thesis work.

Chapter 2

The Genesis Story-Understanding System, circa 2014

The Genesis system understands stories on many levels, as described in detail in a trilogy of papers [Winston, 2011, Winston, 2012a, Winston, 2012b]. Genesis uses the START Parser [Katz, 1997] to read a story written in a simplified subset of English and convert the English to its internal language, Innerese. It then constructs an elaboration graph (Figure 2-1), or visual representation of the story, incorporating common-sense knowledge and concept patterns. Common-sense knowledge includes rules such as "If X kills Y, then Y becomes dead," all specified in plain English. Genesis uses concept patterns to identify concepts such as *revenge*, which can be represented as "X's harming Y leads to Y's harming X." (Figure 2-2.)

In order to model various levels of thinking, Genesis includes multiple versions of some concept patterns. A more sophisticated version of the *revenge* concept incorporating intent is as follows: "X's harming Y leads to Y's wanting to harm X. Y's wanting to harm X leads to Y harming X."

Genesis can interpret stories from different pre-defined cultural points of view, such as Asian or Western. Consider a story in which Eve stabs Bob for no apparent reason. The Asian reader looks for a situational explanation and concludes that Bob must have done something to make Eve angry, whereas the Western reader looks for a psychological explanation and concludes that Eve must be insane. This could be

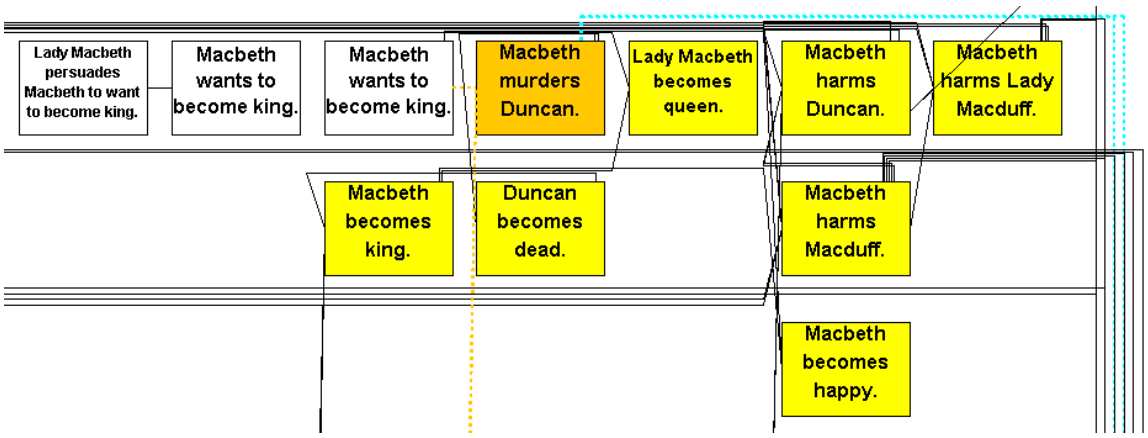

Figure 2-1: This section of the elaboration graph for Shakespeare's *Macbeth* includes events from the story (white boxes) and inferred events (shaded boxes) derived from common-sense rules.

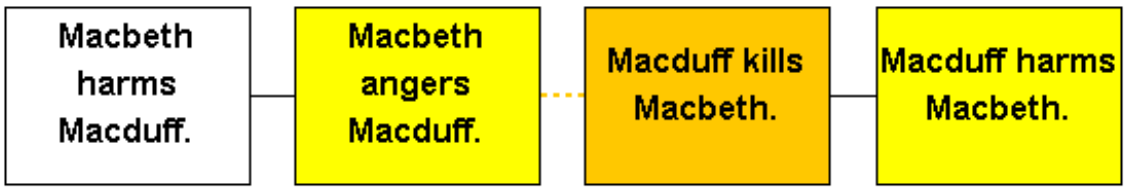

Figure 2-2: In the elaboration graph for Macduff's revenge on Macbeth, Macbeth's harming Macduff leads to Macduff's harming Macbeth.

considered a form of mental modeling, but the points of view need to be pre-defined and there is no way for the model of each character to grow as the story progresses.

Genesis is capable of identifying concept patterns based on instructions to sympathize with a particular character. This is illustrated by the story of the 2007 Russo-Estonia Cyberwar, in which Estonia relocated a war memorial, thereby prompting Russia to attack Estonia's computer networks. From its standard impartial point of view, Genesis recognizes that Russia is taking revenge, whereas from a sympathetic-to-Russia point of view, it believes that Russia is simply teaching Estonia a lesson. In this case, Genesis identifies a different concept pattern—*revenge* or *teaching a lesson*—depending on the point of view (Figures 2-3 and 2-4).

However, this is different from forming a mental model, because concept patterns are hard-coded. Genesis decides which concept pattern to match depending on whether it is told "Russia is my friend." The concept patterns are defined as follows, where (in this case) X is Estonia and Y is Russia:

- *revenge*: "X's harming Y leads to Y's harming X."
- *teaching a lesson*: "Y is my friend. X's angering Y leads to Y's harming X."

The Russo-Estonia Cyberwar analysis is not trivial, but it is a relatively simple matter of matching a pre-defined concept pattern based on the point of view—which country Genesis believes it is a friend of—and does not take into account the actors' perspectives:

- Does Estonia believe it harmed Russia by moving the war memorial?
- Does Russia believe Estonia harmed Russia by moving the war memorial?
- Why does Russia believe it is justified in harming Estonia in response?
- Why does Estonia disagree with Russia?

If Genesis could answer questions such as these, it would be able to understand stories on a much deeper level.

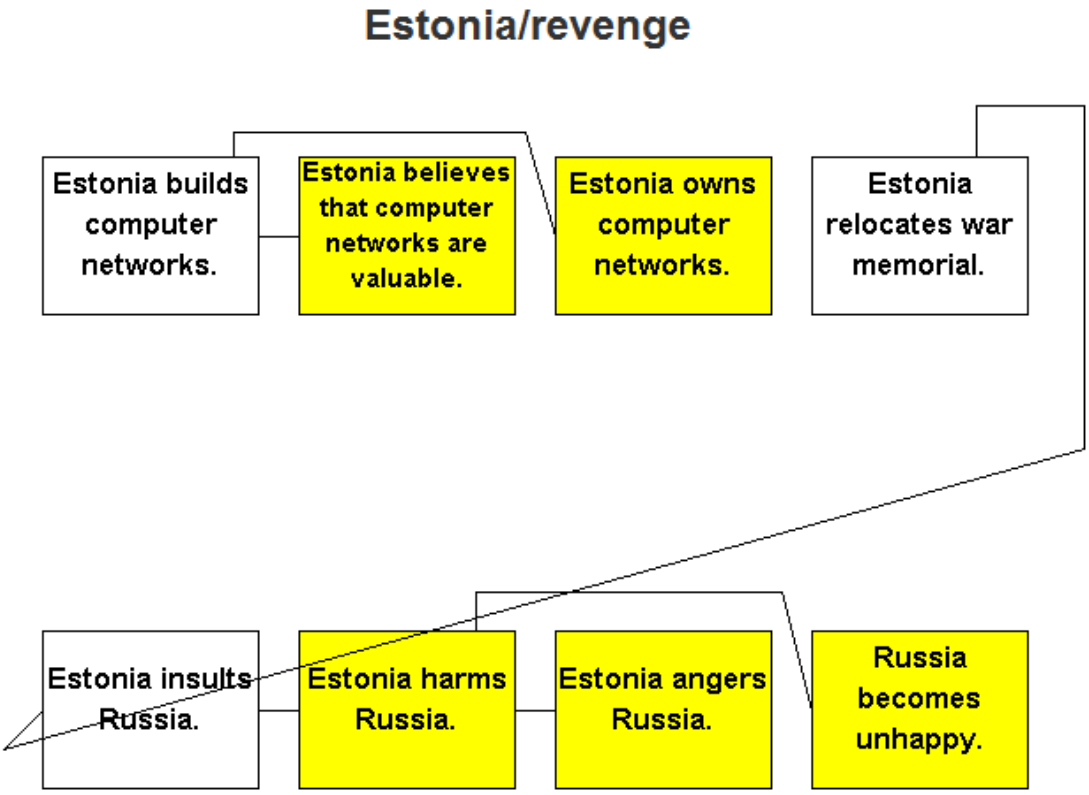

Figure 2-3: Genesis identifies the concept of *revenge* in the Cyberwar story. As before, white boxes represent events that occur in the story, while shaded boxes represent inferences made by Genesis's common-sense rules.

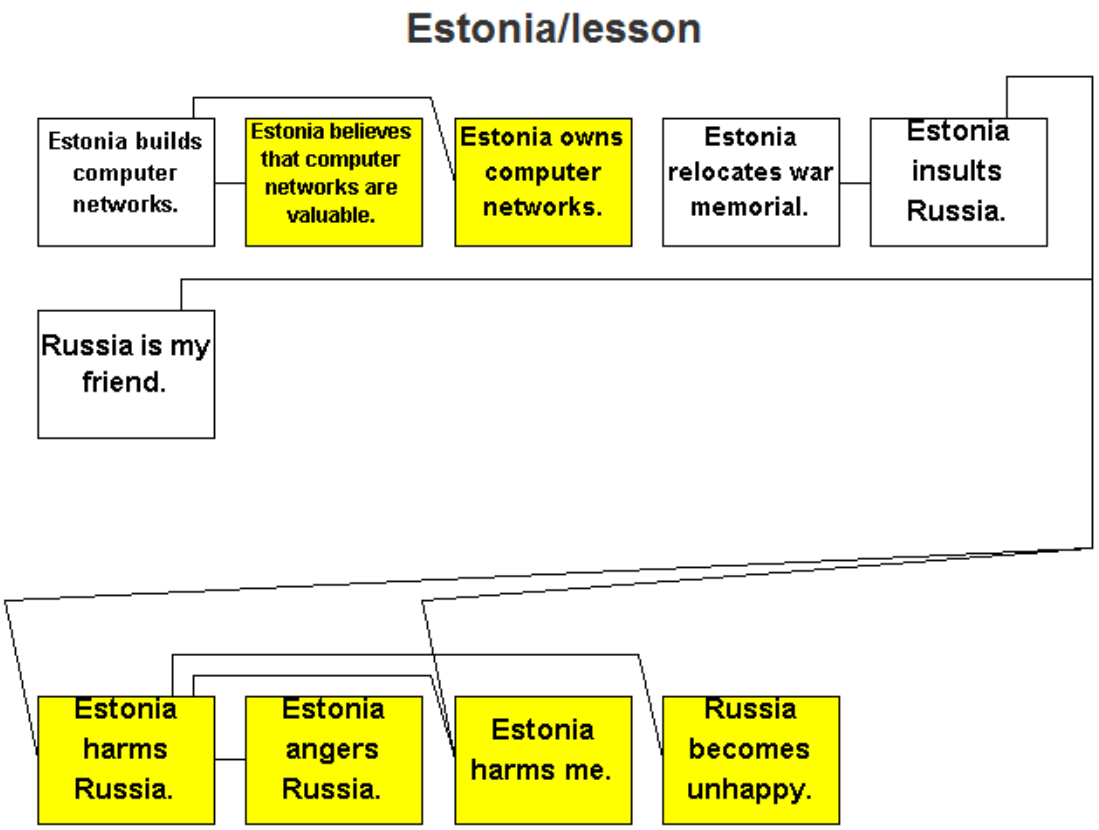

Figure 2-4: When Genesis reads the added line "Russia is my friend," it identifies the concept of *teaching a lesson*.

As of 2014, Genesis could not yet retell a story from a given character's perspective or make a mental model of that character's knowledge of the world, so I undertook to take Genesis to the next level by filling in that hole.

Chapter 3

Forming Mental Models

To enable Genesis to form mental models as it reads a story, I developed the Perspectives Expert, which constructs a mental model of each character's knowledge. Each character's mental model contains the character's personal rules (described in section 3.5), pointers to the events in the original story that the character *observed*, and inferences that the character made using common-sense rules. The story events that appear in a character's mental model are exactly those that the character has *observed*.

Genesis decides which events a character *observes* based on some common-sense assumptions, for example: Person XX observes event EE if and only if XX is present and conscious when EE occurs. By keeping track of each character's observations and inferences, Genesis can construct a mental model of each character's understanding of the world.

In this section, I explain how I first simplified the problem with several rigid assumptions, then gradually made the implementation more robust by adding additional features and removing explicit keyword phrases, to ultimately give Genesis a human-like ability to infer what each character observes and believes.

3.1 Treating stories as simplified theatre makes characters' observations explicit

I started out by annotating a story with entrances and exits, as if it were a play with stage directions. In order to determine whether a character observes an event, the reader needs to know whether that character is present. The script of a theatrical play provides an easy way to know whether a character is present, because it explicitly indicates every scene change and every character's entrances and exits. To simplify the problem of determining whether a character is present, I decided to start by treating non-theatre stories as plays by marking entrances, exits, and scene changes.

As an example, the following story is an extremely abridged summary of the central conflict in Victor Hugo's classic novel *Les Misérables*, written in simplified English, and annotated with entrances, exits, and scene changes:

> Inspector Javert and Jean Valjean enter the scene. Inspector Javert is a policeman. Jean Valjean commits a crime. Inspector Javert and Jean Valjean exit the scene.
>
> Then, Jean Valjean enters the scene. Jean Valjean repents. Everyone exits the scene. A tree falls in the forest.

A human reader would likely notice that there are two scenes (separated by the keyword *Then*), and that Inspector Javert is present only in the first scene, while Jean Valjean is present in both scenes. In the first scene, Jean Valjean becomes a criminal by committing a crime; in the second scene, he repents, becomes a good person, and stops being a criminal. (And then a tree falls in the forest, but nobody is there to hear it.) Thus, a logical interpretation would be that in the end, Jean Valjean knows he is a good person and no longer a criminal, but Inspector Javert still believes that Jean Valjean is a bad person and a criminal because Inspector Javert doesn't know that Jean Valjean has repented. As a result, Inspector Javert will probably still want to capture and imprison Jean Valjean, and Jean Valjean may be confused as to why Inspector Javert would want to imprison a good, non-criminal person who

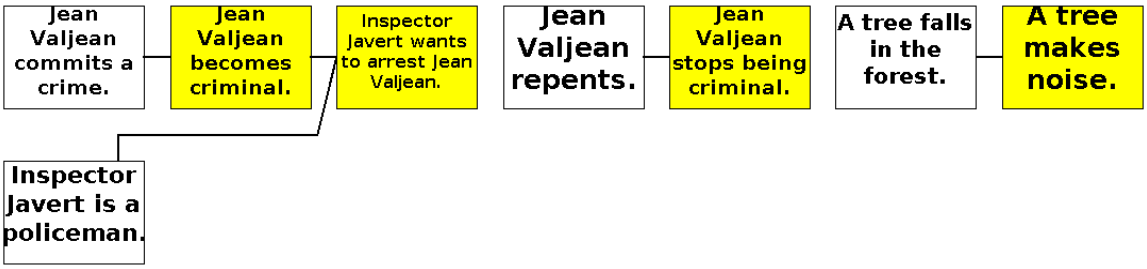

Figure 3-1: Genesis's standard elaboration graph shows all events from the story and all inferred events that can be produced by the provided common-sense rules.

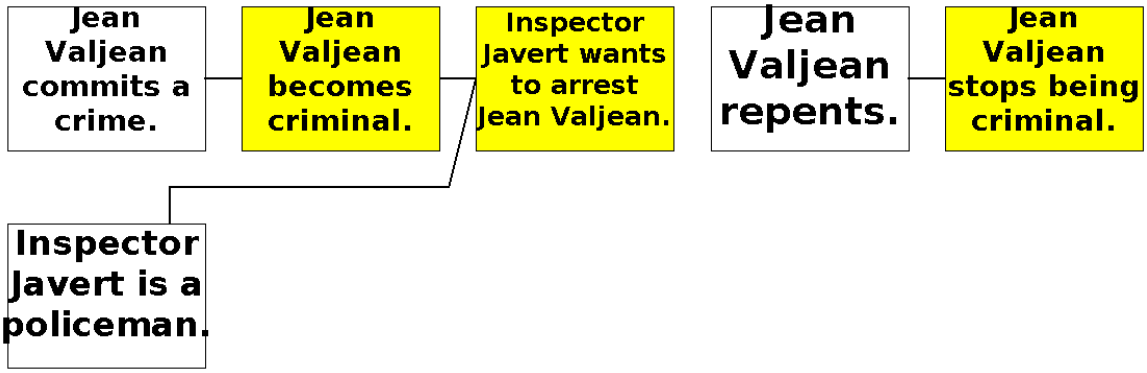

Figure 3-2: The elaboration graph in Jean Valjean's mental model shows the subset events that Jean Valjean observed and the inferences he drew from them.

has repented. This is what a human reader would conclude, so my the goal was for Genesis to reach the same conclusion in the same way.

Figure 3-1 shows the standard elaboration graph that Genesis produces when it reads the story. In my Perspectives Expert I implemented the ability to understand simple stage directions (entrances, exits, and scene changes), resulting in a full mental model for each character in the story. Graphically, each character's mental model is displayed as a personal elaboration graph, as shown in Figure 3-2 (Jean Valjean) and 3-3 (Inspector Javert). As expected, Jean Valjean's mental model includes all the events in the story until the point at which everyone exits, whereas Javert's only includes the events from the first scene, so Javert never stops believing that Jean Valjean is a criminal.

My initial model for determining which events each character observed was to treat stories as simplified theatre: Interpret each story as the script for a play taking place

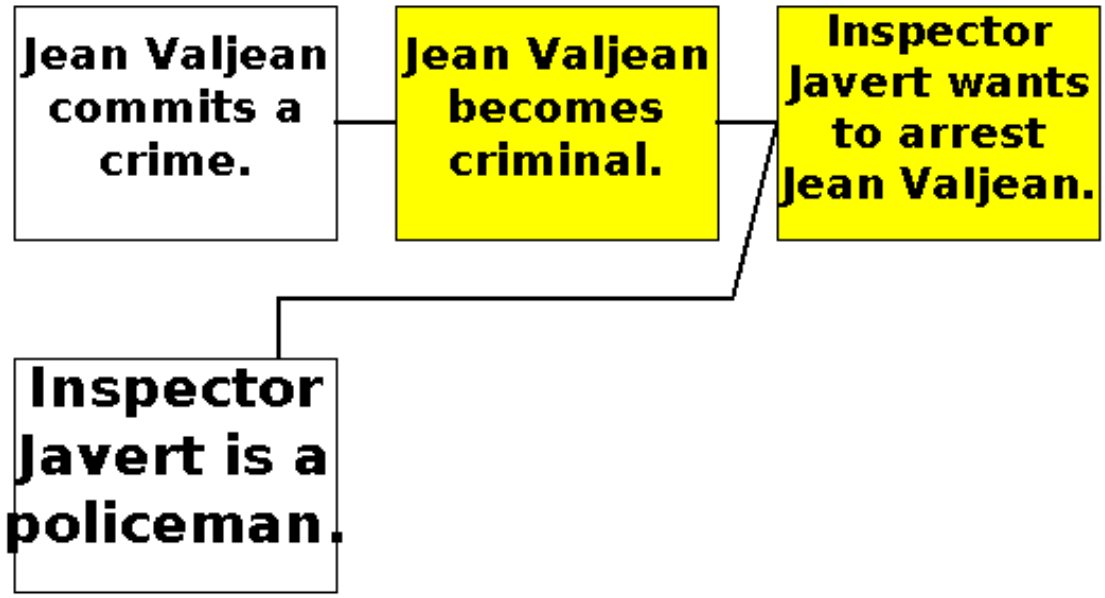

Figure 3-3: The elaboration graph in Inspector Javert's mental model shows the subset of events that Inspector Javert observed—which happen to be a subset of Jean Valjean's observed events—and the events that he inferred.

on a stage. In this oversimplified model, each character is either present onstage, or offstage and not present. Characters have clearly defined entrances and exits, delineated explicitly in the script. If a new scene starts, the previous set of characters exits and a new set (which may or may not include characters who were previously onstage) enters.

Given a few simplifying assumptions, it is reasonable to assume that whenever a character is onstage (and at no other times), they observe all events occurring onstage. The assumptions are:

- There are no monologues or soliloquies while others are physically present.
- All characters onstage are conscious and paying attention.
- Characters who are onstage have no impediments to seeing, hearing, or otherwise observing events occurring onstage. That is, there are no blindfolds, whispering, separate locations depicted in the same scene, etc.
- Characters observe nothing while they are offstage.

- All events take place onstage.

With those assumptions, forming a mental model becomes a simple matter of keeping track of which characters are onstage when each event occurs. I chose to keep track using the paradigm of *opening* and *closing* a character's mental model. When a character enters the scene, their mental model opens; when they exit, it closes. If the scene ends, all mental models get closed. When an event occurs, the event gets recorded as an observation in all mental models that are open at that time. (That is, when a mental model is open, it records observations of any events that may occur.)

Each character's mental model contains a story processor, similar to the main Genesis story processor. Each character's story processor uses that character's personal observations to reach conclusions based on global common-sense rules, such as "If XX is hungry, then XX wants to eat" or "If XX commits a crime, then XX becomes a criminal." If the antecedent (the IF part) of a rule matches the observed or inferred events in a character's mental model, then the rule fires and the bound consequent (the THEN part, with variables replaced by names) gets added to their mental model.

Several keyword phrases help Genesis to detect entrances, exits, and scenes changes. The idioms "XX enters the scene" and "XX exits the scene" indicate that Genesis should open or close the character's mental model, respectively. Similarly, the idiom "Everyone exits the scene" indicates that Genesis should close all currently open mental models. Finally, the keyword *Then* acts as a scene marker, indicating that the previous scene has ended and a new scene is starting. When Genesis sees the keyword *Then*, it closes all currently open mental models. Thus, the example story above could equivalently be rewritten: The sentence "Inspector Javert and Jean Valjean exit the scene" could be replaced with "Everyone exits the scene," or it could even eliminated completely because Genesis will close all mental models before beginning the second scene.

The process of determining which events each character observed became straightforward after I annotated stories with stage directions, treating them as simplified

theatre. The next step would be for Genesis to determine characters' observations *without* stage directions.

3.2 Inferring observations from context eliminates stage-direction annotations

Even children's stories do not explicitly indicate characters' entrances and exits, yet children can easily infer which characters are present at most points in a story. To eliminate the explicit entrances and exits, I started by making the assumption that all characters who are mentioned in a scene (as the subject of a sentence) are present throughout the entire scene. This assumption is naïve in the sense that Genesis may get false positives when a character is mentioned in dialogue, or in a sentence such as "David was not present," but for the most part, it seems to be a reasonable assumption.

Here is the same simple story from above, without the explicit entrances and exits, and this time split up into three scenes. As before, the keyword *Then* introduces a new scene.

> Inspector Javert is a policeman. Jean Valjean commits a crime.
>
> Then, Jean Valjean repents.
>
> Then, a tree falls in the forest.

In addition to all the assumptions from the previous section, I added the assumption that if character is mentioned, it means they were present since the scene began, and thus they observe everything that happens from the beginning of the scene until they exit or the scene ends. At this level, the only exception to the assumption that a character is present from the beginning of a scene is if they are mentioned in an entrance statement, e.g. "Emily enters the scene." In that case, Genesis infers that the character enters as indicated, as opposed to assuming that the character was present all along.

When Genesis reads this revised story, with no stage directions, it still produces the same elaboration graphs: Jean Valjean observes events in the first two scenes (Figure 3-2), while Inspector Javert only observes the first scene (Figure 3-3).

When Genesis detects a character's name, it first checks whether that character's mental model is currently open. If so, nothing changes. On the other hand, if the character does not even have a mental model yet, Genesis creates one, with the mental model starting out closed by default. Finally, if the character's mental model is closed, Genesis will open it, but here is the tricky part: In order to capture all events that occurred before the character's name was mentioned, Genesis rewinds to the beginning of the scene (denoted by the scene marker *Then*, or by the beginning of the story), inserts the character's entrance (e.g. "Peter enters the scene") at the beginning of the scene, then rereads the scene line by line so that each event gets correctly inserted into the character's mental model (which is now open). There is no concern about accumulating duplicate events in other characters' mental models as Genesis rereads, because Genesis always consolidates identical events within a scene. (For example, in the story "Bobby jumped. Bobby jumped. Bobby jumped.", Bobby only jumps once, as far as Genesis is concerned.)

At first glance, this procedure of rewinding and rereading a scene in the story—which may occur multiple times—may seem to have two glaring issues: Genesis could get stuck in an infinite loop, and the procedure appears wildly inefficient. Consider the following single-scene example story which would result in an infinite loop:

> Bobby bakes a cake. Alice gets bored. Alice exits the scene. Bobby watches the oven. Alice smells the cake. Bobby takes the cake out of the oven.

In reading the story, Genesis would take the following steps:

1. Read "Bobby bakes a cake." Rewind to insert "Bobby enters the scene" at the beginning of the story.

2. Read "Bobby enters the scene" and open Bobby's mental model.

3. Reread "Bobby bakes a cake" and record it in Bobby's mental model.

4. Read "Alice gets bored." Rewind to insert "Alice enters the scene" at the beginning of the story.

5. Read "Alice enters the scene" and open Alice's mental model.

6. Reread "Bobby enters the scene" and "Bobby bakes a cake"; record "Bobby bakes a cake" in Alice's mental model.

7. Reread "Alice gets bored" and record it in Alice's and Bobby's mental models.

8. Read "Alice exits the scene" and close Alice's mental model.

9. Read "Bobby watches the oven" and record it in Bobby's mental model (because Bobby's is still open, but Alice's is now closed).

10. Read "Alice smells the cake." Alice's mental model is closed, so rewind to the beginning of the scene to insert Alice's entrance.

11. Reread "Bobby bakes a cake. Alice gets bored. Alice exits the scene." Close Alice's mental model.

12. Reread "Bobby watches the oven." (Bobby is mentioned, but his mental model is already open, so no need to do anything special.)

13. Reread "Alice smells the cake." Uh-oh! Alice's mental model is closed, so rewind to the beginning of the scene to insert Alice's entrance... and now Genesis is stuck in an infinite loop, and Bobby will never take the cake out of the oven.

At first glance, this may look like a huge oversight. But on closer examination, what should we expect Genesis to do? Alice explicitly exited the scene, and then later she smells the cake. Either she's smelling the cake from offstage, which violates the assumption that only characters onstage can observe events, or she has re-entered... except that we have no way of determining at which point she re-entered after her explicit exit. Thus, I believe Genesis is justified in not knowing what to do in this

ill-defined story, and Genesis's behavior of getting stuck in an infinite loop is really just way of saying "Help! This story is ambiguous! What do I do?!" In summary, the mechanism of rewinding and rereading is not infallible, but neither are human readers when presented with an ambiguous story.

As for the apparent inefficiency of repeatedly rewinding and rereading, I believe it's not so far from what a human would do if presented with the task of forming a complete mental model for each character. They would start off reading the sentences of the story in order, one by one. They would have no way of predicting when a new character might appear, so when a new character is suddenly mentioned and seems to have been present the entire time, the reader would need to have some mechanism for going back to figure out which events the new character must have observed. As humans, we may feel that this recall is instantaneous, but whether we are rewinding and replaying very quickly, or effectively copy-and-pasting from our own mental model, the time required is non-zero and would probably scale with the length of the story and the number of characters introduced.

Consider the following scene:

> Alice suggests, "Maybe we could go for a bicycle ride next weekend." Bobby replies, "No, we should go camping instead." Alice complains, "But I want to ride my new bike." Bobby retorts, "Yes, but I just got a new lightweight tent." Alice and Bobby continue arguing. After sitting silently through the conversation, Carla interjects, "Why not do both at once?"

Even though Carla is not mentioned until the end of the scene, we can easily infer that Carla was there the whole time, and that she knows that the conversation is an argument about bicycling and camping. But how does the reader know exactly which statements Carla observed? Unless the reader has a perfect memory of everything that Alice and Bobby said, as well as when the scene started, they would need to either rewind and replay quickly, or copy and paste from their mental model of Alice or Bobby.

It is also worth noting that human readers often have no reason to form complete

mental models of all characters. As my copy editor, Rebecca Kekelishvili, pointed out, we often do not even bother consciously forming mental models at all: "When I read, and I see a new character enter, sometimes, I do not wonder about what they know. I just take them for granted and move on with the book." Genesis is a much more careful reader than many humans, and it may even discover conclusions that a human might miss. Genesis's reading of a story resembles that of a student whose teacher stops them after each sentence and asks, "And what new information does that sentence tell you?"

By keeping track of which characters are mentioned in each scene, I enabled Genesis's Perspectives Expert to identify each character's observations without stage directions indicating the character's entrances and exits. Next, I wanted to eliminate the stage aspect of my simplified theatre model by supporting multiple locations, instead of requiring all scenes to take place in series on the same stage.

3.3 Locations help the reader keep track of characters

Thus far, I have represented a story with a single stage upon which all events take place, and which characters can enter and exit. If a story switches between locations—for example, alternating between scenes in the house and scenes in the office—the scene must be reset each time, and Genesis needs to be told which characters re-enter. However, when humans read stories, we typically remember which characters were present in any particular location, so the story can switch between established locations without needing to explicitly tell the reader who is present each time. In this section, I describe how I enabled Genesis to replicated this human behavior by keeping track of characters' locations, so that Genesis can infer which characters are present based on known locations.

I started by replacing the scene markers (*Then*) with locations, as demonstrated in the example story below (loosely inspired by *Les Misérables*). A chapter in a novel

often begins with a long description of a new location; the keyword phrase "Scene shifts to [location]" can be thought of as a simplified English summary of such a paragraph.

> Scene shifts to the bakery. Inspector Javert and Jean Valjean enter the bakery. Inspector Javert is a policeman. Jean Valjean commits a crime. Jean Valjean exits the bakery. Jean Valjean enters the prison. Inspector Javert laughs. Inspector Javert exits the bakery.
>
> Scene shifts to the prison. Inspector Javert enters the prison. Inspector Javert punishes the inmates.
>
> Scene shifts to the forest. Jean Valjean exits the prison. Jean Valjean enters the house. A tree falls in the forest.
>
> Scene shifts to the prison. The inmates work.
>
> Scene shifts to the house. Jean Valjean repents.

The full Genesis elaboration graph is shown in Figure 3-4, and the two characters' mental models are shown in Figure 3-5 (Jean Valjean) and Figure 3-6 (Inspector Javert). Note that each character observes events if and only if they are present in location of the event when the event occurs, and when the story returns to a previous location (the prison), the Perspectives Expert remembers which characters are still present in that location (Inspector Javert). Specifically, it knows that Jean Valjean has left the prison, and that Inspector Javert has not.

In addition to being more realistic, one advantage of locations over scenes is that there can be parallel storylines, and the reader (in this case, Genesis) can keep track of the characters in each location. To keep track of characters, I introduced the concept of a *container*, as a data structure within the Perspectives Expert. Each container represents a location and can contain multiple characters. A character can be in multiple containers at once, for instance in the kitchen and on the phone.

More generally, containers provide scope for what characters can perceive. In principle, in addition to containers representing different locations, there could be

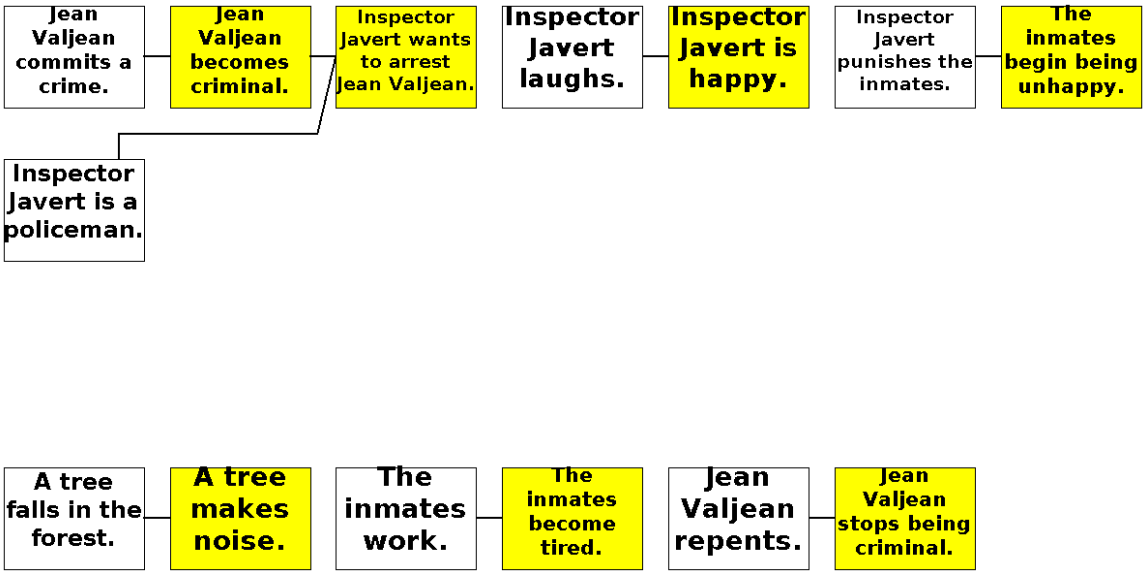

Figure 3-4: Genesis's main story processor ignores locations and constructs a standard elaboration graph.

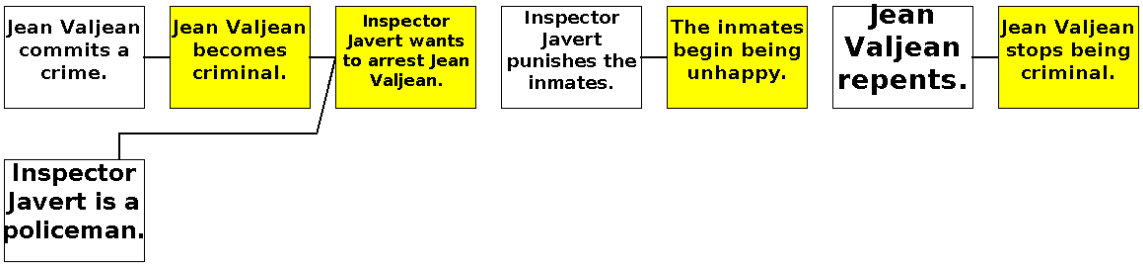

Figure 3-5: Jean Valjean's elaboration graph includes the events that occurred in the same temporospatial locations that he was in. This includes the events that occurred in the bakery before Jean Valjean exited, the events that occurred in the first prison scene, and Jean Valjean's repentance in the house.

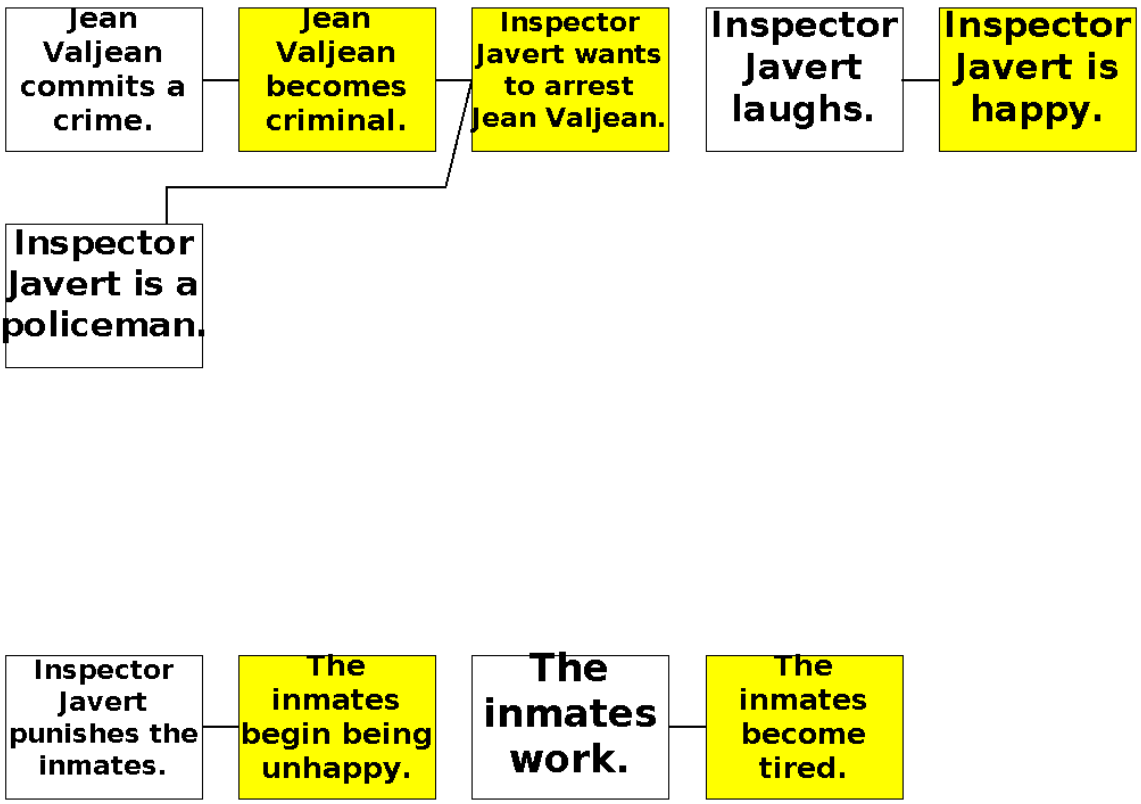

Figure 3-6: Inspector Javert's elaboration graph includes the events that occurred in the same temporospatial locations that he was in—specifically, all events that occurred in the bakery and in the prison, even though Inspector Javert was not mentioned in the second prison scene.

separate containers for different senses such as sight and smell, and there could even be containers nested within other containers, representing containment, or one space being within another (e.g. the kitchen is within the house, and the house is within the neighborhood). Thus, containers can capture the fact that all the people in a room can hear each other, that two people in the phone can hear both sides of the conversation but eavesdroppers can't, and that you can see food through a window but can't smell it.

Containment gets complicated quickly because it necessitates some notion of size and distance: If XX and YY are in the same kitchen, they can probably observe all of each others' actions, but they are unlikely to be within eyesight or hearing distance if they are simply in the same city. As a first step, I started by implementing containers for locations only.

To keep track of each character's location(s), I created a hashmap in the Perspectives Expert, mapping each location (*container*) to a list of the characters in that location. When a character enters a location, I add them to the appropriate container in the hashmap, and if the current scene is set in that location, then I also open their mental model. For example, if Emily enters the kitchen, and the current scene is taking place in the kitchen, then the Perspectives Expert would add Emily to the "kitchen" container and open her mental model. Similarly, when a character exits a location, the Perspectives Expert removes them from that container in the hashmap. If that container was the current scene location, it closes their mental model. In addition to containers defined by explicit locations (the kitchen, the bedroom, etc.), there is a default container in which the scene takes place if no location is explicitly mentioned in the story. (This default container also serves the purpose of maintaining backward compatibility with stories written in the simplified-theatre style using idioms such as "XX enters the scene.")

When the scene shifts to a different location—marked by the idiom "Scene shifts to [location]"—Genesis closes the mental models of all characters from the previous location. Then, Genesis opens the mental models of any characters known to be in the new location.

The hashmap keeps track of all of the characters in every location, so that Genesis effectively remembers which characters were last present in each location. Thus, if the scene shifts back to a previous location ("Meanwhile, the argument in the kitchen was still ongoing..."), Genesis can easily reopen the mental models of everyone who was in that location.

With this container representation, each location has a container, and each character can be in zero or more containers. An equivalent representation that I proposed is *awareness windows*: Each character could have an awareness window, representing the locations they are able to observe, and each awareness window would contain zero or more locations.

However, I believe that containers are more similar to human behavior because we tend to store memories of who was in each location, and we recall a memory when we return to a specific location, whether in real life or while reading a story. This is consistent with Marvin Minsky's K-lines theory of memory [Minsky, 1979], which says that we remember by putting ourselves back in a partial mental state corresponding to the original memory. When a human reader sees the phrase "Meanwhile, in the town of Montfermeil...", they activate the K-line corresponding to the town of Montfermeil to remember what was happening there—and which characters were present—when they last heard about the town.

By introducing *containers* as a means of keeping track of characters' locations in the Perspectives Expert, I enabled Genesis to understand multiple-location stories without changing the mental model framework.

3.4 Unconscious characters cannot observe events

If a character is unconscious—whether they are asleep, knocked out, dead, or simply daydreaming and not paying attention—they clearly can't observe most events in a story. Of course, there are edge cases, such as the sleeping person who is aware of a real-world loud noise but thinks it's just part of their dream. But for the most part, unconscious characters do not observe anything.

Accordingly, I enabled Genesis to detect unconsciousness so that it only records events in a character's mental model when the character is conscious. For example, in the following story, we expect that Marius will observe his friend Enjolras's death, but Marius will not observe himself being rescued by Jean Valjean because he is unconscious throughout the rescue.

> At the barricade, Marius fights in the uprising with Enjolras. Enjolras gets injured. Enjolras dies. Marius gets injured. Marius loses consciousness. Jean Valjean carries Marius into the sewer toward safety. Jean Valjean rests. Marius is covered in mud. Jean Valjean falls asleep. Then, Thenardier steals valuable ring from Marius. Jean Valjean snores. Then, Jean Valjean awakens. Marius is covered in filth. Thenardier exits the scene. Jean Valjean continues carrying Marius. Jean Valjean leaves Marius on Grandfather's doorstep. Jean Valjean exits the scene. Then, Marius regains consciousness. Marius is grateful. Marius wants to know the identity of his rescuer.

Genesis's standard elaboration graph is shown in Figure 3-7. In order to make the story easier to understand (for human readers), here is the same story, annotated with commentary in italics and with indentations to better visualize which characters are awake at each point in time:

> At the barricade, Marius fights in the uprising with Enjolras.
>
> Enjolras gets injured.
>
> > Enjolras dies. *Enjolras stops observing.*
> >
> > Marius gets injured.
> >
> > > Marius loses consciousness. *Marius stops observing.*
> > >
> > > Jean Valjean carries Marius into the sewer toward safety.
> > >
> > > Jean Valjean rests.
> > >
> > > Marius is covered in mud.

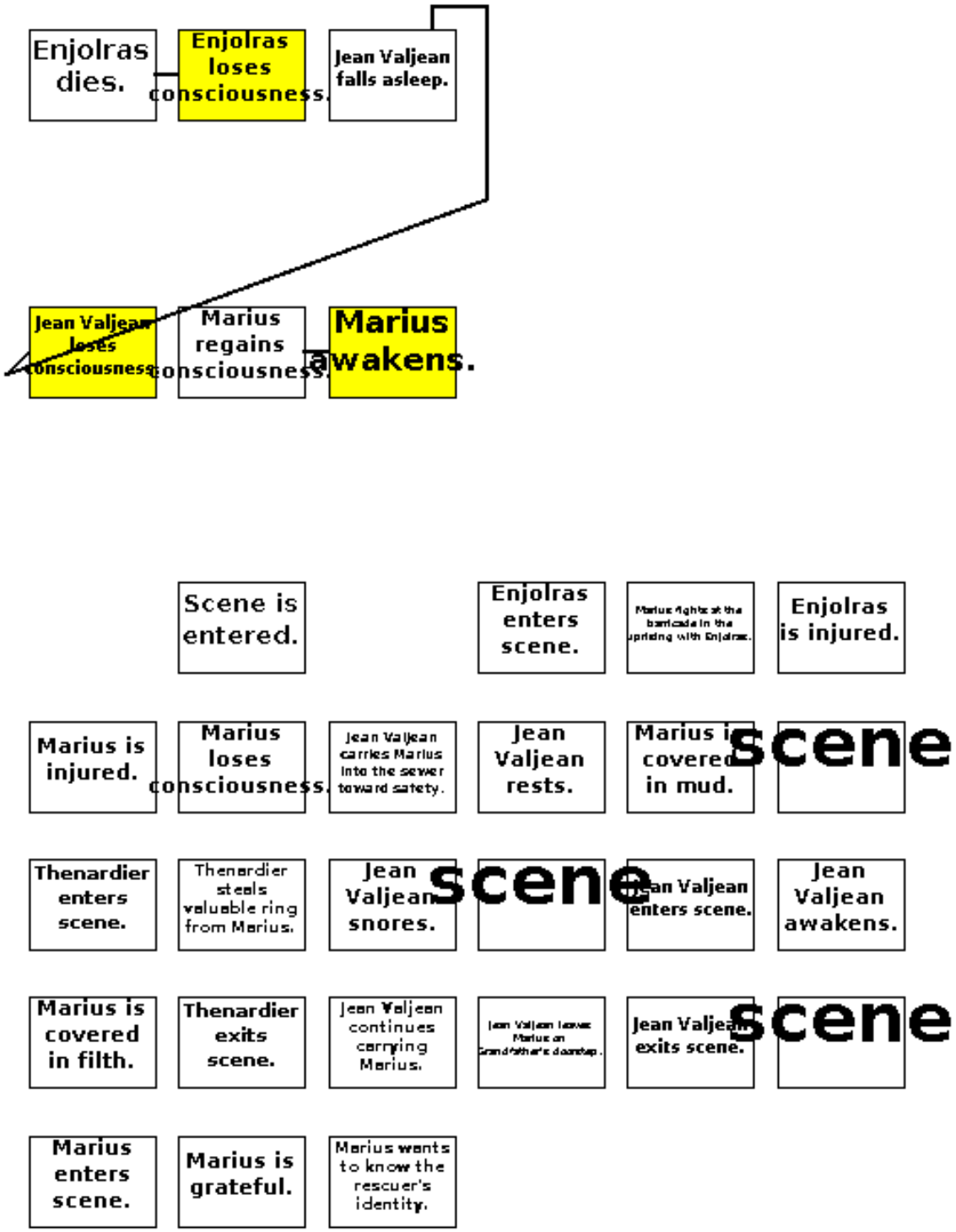

Figure 3-7: Genesis's elaboration graph for the unconsciousness example story shows all events in the story.

Jean Valjean falls asleep. *Jean Valjean stops observing.*

New scene begins. Thenardier and Jean Valjean are present, but Jean Valjean still is not observing.

Then, Thenardier steals valuable ring from Marius.

Jean Valjean snores.

New scene begins. Thenardier, Jean Valjean, and Marius are present, but Marius remains unconscious, and Thenardier exits mid-scene.

Then, Jean Valjean awakens. *Jean Valjean resumes observing.*

Marius is covered in filth.

Thenardier exits the scene.

Jean Valjean continues carrying Marius.

Jean Valjean leaves Marius on Grandfather's doorstep.

New scene begins. Only Marius is present.

Then, Marius regains consciousness. *Marius resumes observing.*

Marius is grateful.

Marius wants to know the identity of his rescuer.

When Genesis reads the (unannotated) story above, it correctly forms mental models of each character based on their presence and consciousness (or lack thereof) at each point in the story. Figures 3-8, 3-9, 3-10, and 3-11 show the elaboration graphs for Enjolras, Marius, Jean Valjean, and Thenardier, respectively.

By default, all characters are assumed to be conscious. This makes sense, because as a human reading a story, you would usually assume that if Jenny is introduced in a scene, she is probably awake and conscious unless the story explicitly states otherwise.

When Genesis sees the phrase "XX loses consciousness," where XX is a character's

Figure 3-8: Enjolras's elaboration graph includes only the events that happened before Enjolras's death. He is not aware of his own death.

Figure 3-9: Marius's elaboration graph includes the events that occur before he loses consciousness and after he regains consciousness. Marius observes his friend Enjolras's death, then he becomes injured and loses consciousness. His next memory is of regaining consciousness. He is grateful to be alive, but does not know who rescued him.

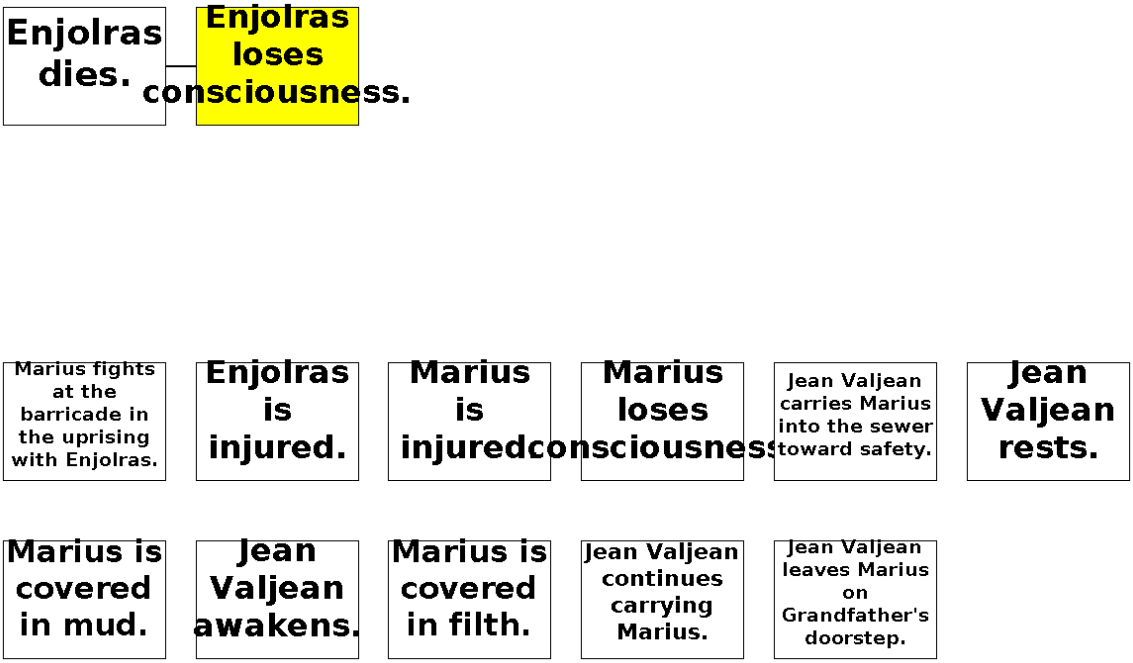

Figure 3-10: Jean Valjean's elaboration graph includes most of the events in the story, except those that occurred while he was asleep (Thenardier stealing Marius's ring) and after Jean Valjean exited (Marius awakening).

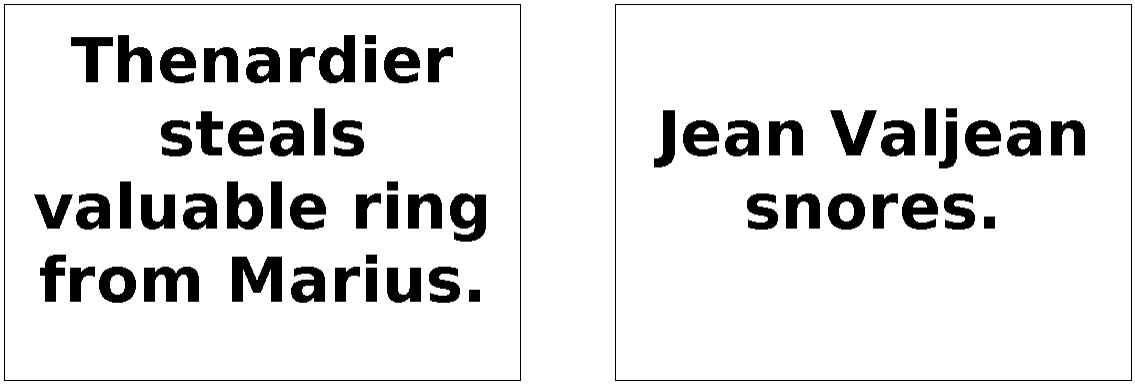

Figure 3-11: Thenardier never loses consciousness in this story, so his elaboration graph simply includes the events that occurred while he was present.

name, it closes XX's mental model—signifying that XX has stopped observing—and adds XX to a list of unconscious characters. Similarly, when it sees "XX awakens", it removes XX from the list of unconscious characters. However, Genesis only opens XX's mental model if XX is also present in the current scene: If Frank wakes up in the bedroom but the story is currently describing events in the kitchen, there is no reason to open Frank's mental model until either the scene shifts to the bedroom or Frank enters the kitchen.

As far as the Genesis codebase is concerned, there is only one type of losing consciousness ("loses consciousness") and one type of regaining consciousness ("awakens"). The variations on "XX loses consciousness" and "XX awakens" are handled by deduction rules such as "If XX falls asleep, then XX loses consciousness" that convert alternate phrases into the keyword phrases that Genesis recognizes.

In terms of efficiency and human plausibility, maintaining a list of unconscious characters only makes sense if the majority of characters in the story are always awake, and if those who become unconscious or regain consciousness do so explicitly. For example, it would be very inefficient to use Genesis's one-character-at-a-time implementation in a story describing 1000 sleeping characters who take turns awakening briefly one at a time, for which it would make more sense to maintain a list of awake characters. However, it seems reasonable to assume that most stories describe conscious characters who are paying attention most of the time, in which case Genesis's method is certainly plausible as a model of human behavior.

3.5 Personal Rules: Individuals use different assumptions about the world

In most stories—and in real life—every person has different assumptions about the world. Previously, Genesis only supported a single set of global rules (Genesis's assumptions about the world), so I added the ability for each character to make inferences using their own set of personal rules.

Thus far, we have assumed that Jean Valjean and Inspector Javert's difference of opinion is a result of Javert not seeing Valjean's repentance. However, another common interpretation of the story is that Javert was aware of everything that Valjean did, but Javert had a different set of rules governing how the world works. In the following example story, there are no global rules about repentance, as is evident in Genesis's elaboration graph of the story (Figure 3-12). Instead, the two characters have personal rules: Jean Valjean believes that repenting can make someone stop being a criminal, while Javert believes that a criminal who appears to have repented is actually just a liar, so he refuses to believe that Valjean has stopped being a criminal.

Global common-sense rules:

If XX commits a crime, then XX becomes criminal.

If XX is a policeman and YY becomes criminal, then XX wants to arrest YY.

If XX stops being criminal, then XX will follow the law.

Story:

Inspector Javert thinks like personality-p.

Jean Valjean thinks like personality-j.

Inspector Javert is a policeman. Jean Valjean commits a crime. Jean Valjean repents.

In the main story file, when Genesis sees the idiom "[name] thinks like [personality type]"—for example, "Jean Valjean thinks like personality-j"—it imports all the rules from the corresponding personality file into the named character's mental model. For example, the `file personality-j.txt` defines the personality traits of someone with personality-j (in this case, Jean Valjean). Each personality file contains a collection of common-sense rules, similar to the global common-sense rules in a story file. After importing a character's personal rules into their mental model, Genesis uses the personal rules to draw inferences from the perspective of the character. The inferences

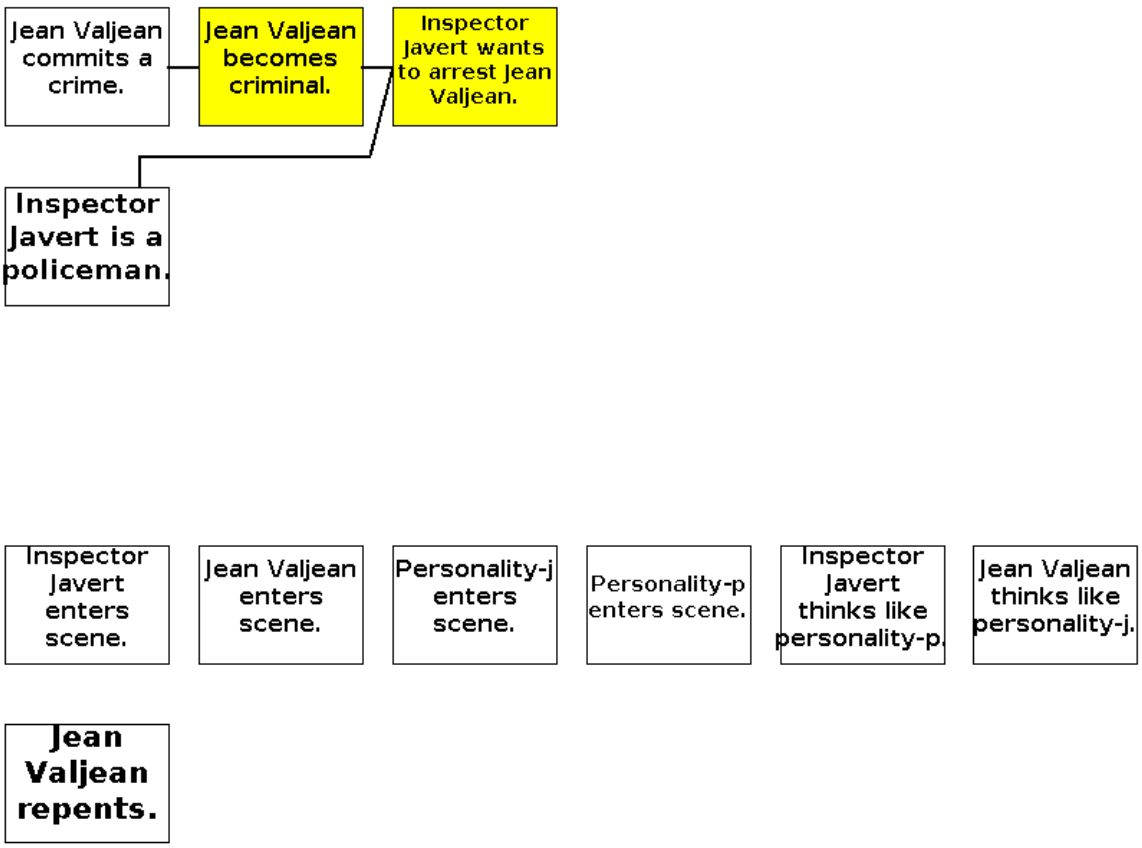

Figure 3-12: Genesis's standard elaboration graph includes only the events in the story and those inferred from global rules. In this story, Genesis has no global rules about repentance.

then appear in the character's elaboration graph just like inferences made by global rules. (In my current model, characters are not aware of each other's personal rules, but that would be an interesting next step toward a higher level of mental modeling.)

For this example story, Inspector Javert's personality file (`personality-p.txt`) contains his three personal rules:

If XX becomes criminal, then XX should go to prison.

If XX becomes criminal and XX repents, then XX is a liar.

If XX commits a crime, then XX must be a bad person.

Similarly, Jean Valjean's personality file (`personality-j.txt`) contains his two personal rules:

If XX repents, then XX stops being criminal.

If XX commits a crime, then XX must be poor and desperate.

In addition to the characters' difference of opinion over what happens when Jean Valjean repents, Genesis also notices that the two characters have different explanations for why Jean Valjean committed a crime in the first place. These explanations are given by *explanation rules*, a standard Genesis rule type that uses the phrase "must be" in the consequent (the THEN clause). As shown in the characters' mental models, Inspector Javert believes that all people who commit crimes are bad (Figure 3-13), whereas Jean Valjean believes that people are sometimes driven to steal because they are simply poor and desperate and have no alternative (Figure 3-14).

3.6 Real stories have complex scenes: Partial information and additional complications

One of the biggest challenges of modeling human story understanding is that there are always exceptions. In general, when a character becomes dead, they stay dead... but once in a while they come back to life in a metaphysical state, or it turns out

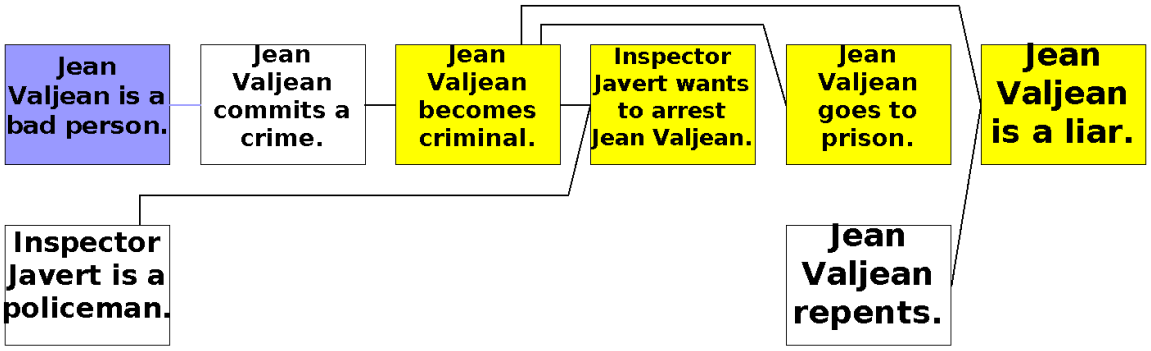

Figure 3-13: Inspector Javert's elaboration graph incorporates his personal rules, which say that someone who commits a crime must be a bad person, criminals should go to prison, and any criminal who appears to have repented is a liar.

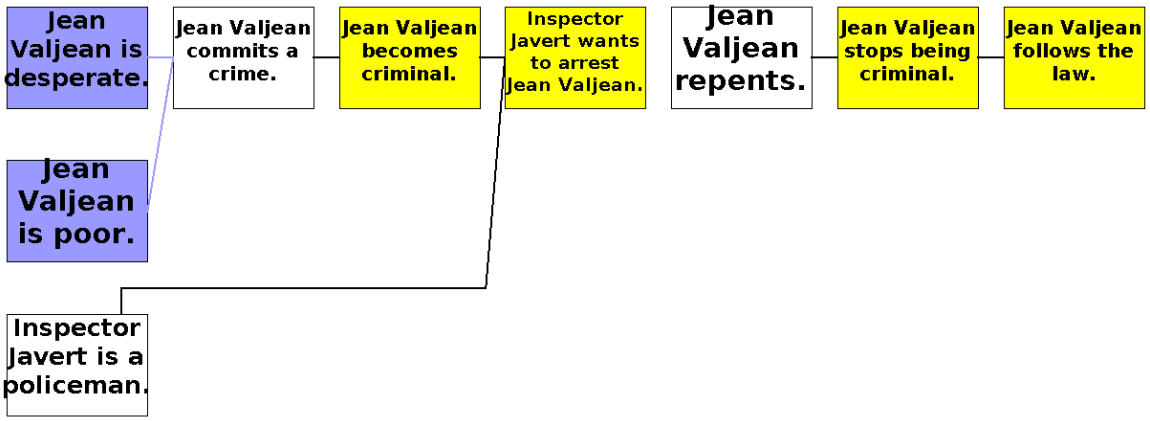

Figure 3-14: Jean Valjean's elaboration graph incorporates his personal rules, which say that a person who repents will stop being a criminal. This invokes Genesis's global rule that if someone stops being a criminal, then they will follow the law. Jean Valjean also believes that people commit crimes not because they are bad, but because they are poor and desperate.

they weren't dead to begin with. In general, when two awake people are standing next to each other and the first person makes a noise, the second person can hear it... unless the second person is deaf or wearing earplugs. Accordingly, I discovered that no matter how many special cases I covered, there were always additional ones that my model did not address. The following is a list of some of the more significant complications that I discovered through my research but did not address:

- **Partial information due to limited senses**: If some observation channels (hearing, seeing, etc) are disabled, a character will only have partial information about a scene. For example, if Alice is looking in through Bobby's kitchen window, she may be able to see him cooking but not be able to smell that the food is burning. Or, if Alice is sitting in the kitchen but blindfolded, then she will be able to smell the food and hear it sizzling, but she won't be able to see it.

- **One-way information**: Sometimes one character can observe the other, but not vice versa. For example, if David is using binoculars to see some people on a distant hill, he may be able to see everything the people are doing, but the people are unlikely to see him. Similarly, if Emily and Frank are talking to each other on the phone, and Carla is sitting next to Emily, then Carla can probably hear Emily's side of the phone call but not Frank's, whereas Frank may be able to hear if Carla speaks.

- **Inability to understand**: If one group of characters are speaking in different languages, communicating in code, or referencing inside jokes, then others who are not in that group may not be able to understand.

- **Forgetfulness**: Just because a character has observed an event does not necessarily mean they will remember it forever.

Chapter 4

Applications of Mental Models and Perspectives

4.1 Retelling the story from a character's perspective

Why retell stories? One of the easiest ways to determine whether someone has formed an accurate mental model of a character is to ask them to retell the story from that character's point of view. This is a task that young children (under age 4) have trouble with, because they do not yet have the ability to differentiate their own personal knowledge from what other people know—that is, they do not yet have the ability to form mental models of other people's thoughts, as demonstrated in several False Belief experiments [Call and Tomasello, 1999].

Because retelling a story is a good demonstration of understanding, one of my early goals for Genesis's mental-modeling capability was for Genesis to be able to retell a story from the perspective of any character in the story.

By way of demonstration, here again is a very simple version of *Les Misérables*:

> Inspector Javert is a policeman. Jean Valjean commits a crime.
>
> Then, Jean Valjean repents. Jean Valjean becomes a good person.

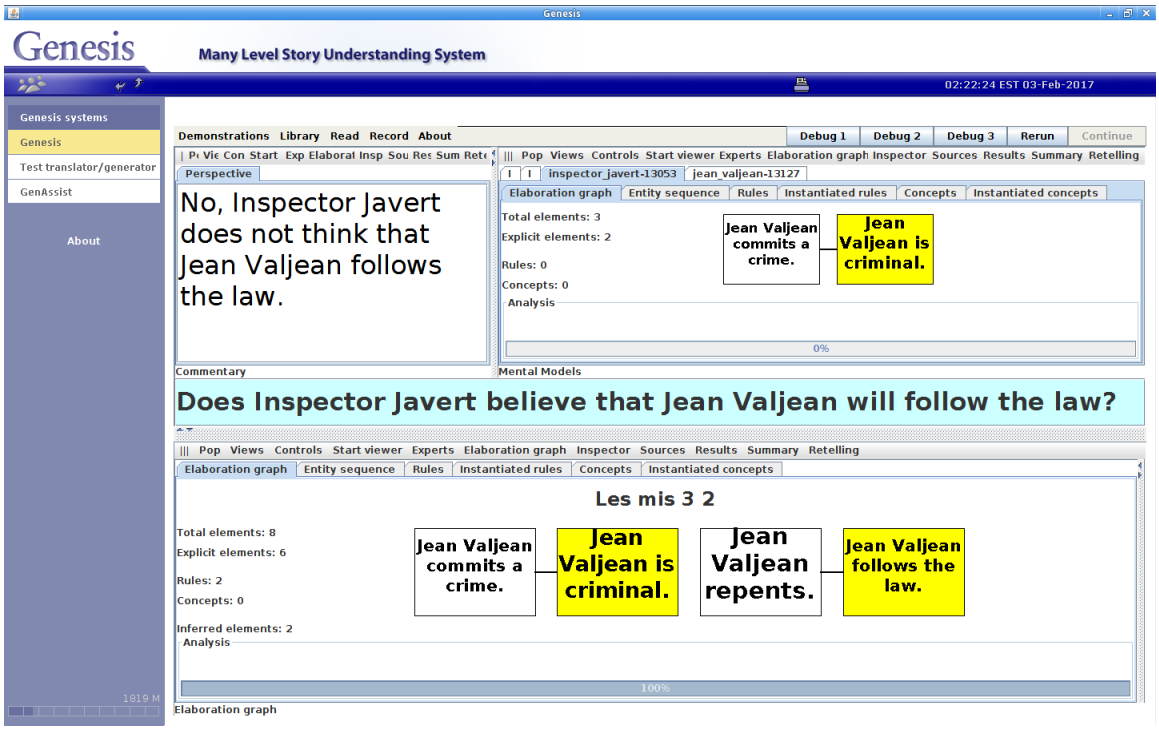

Figure 4-1: Genesis has a text box in which users can ask questions and enter commands.

To allow users to interact with Genesis, the graphical interface has a text box in which users can ask questions and enter commands (Figure 4-1).

If the user enters "Tell story as Inspector Javert," then Genesis will output:

> **Inspector Javert's story:**
>
> Inspector Javert is a policeman. Inspector Javert wants to catch criminals because Inspector Javert is a policeman. Inspector Javert wants to catch criminals. Jean Valjean commits a crime. Jean Valjean is criminal because Jean Valjean commits a crime. Jean Valjean is criminal. Inspector Javert wants to arrest Jean Valjean because Jean Valjean is criminal, and Inspector Javert is a policeman. Inspector Javert wants to arrest Jean Valjean.

Similarly, if the user enters "Tell story as Jean Valjean," Genesis produces Jean Valjean's version of the story (with changes in bold):

Jean Valjean's story:

Inspector Javert is a policeman. Inspector Javert wants to catch criminals because Inspector Javert is a policeman. Inspector Javert wants to catch criminals. Jean Valjean commits a crime. Jean Valjean is criminal because Jean Valjean commits a crime. Jean Valjean is criminal. Inspector Javert wants to arrest Jean Valjean because Jean Valjean is criminal, and Inspector Javert is a policeman. Inspector Javert wants to arrest Jean Valjean. **Jean Valjean repents. Jean Valjean becomes a good person. Jean Valjean isn't criminal because Jean Valjean becomes a good person, and Jean Valjean repents. Jean Valjean isn't criminal.**

With Genesis's mental-model representation, retelling a story becomes a surprisingly trivial application of mental models, because each character's mental model already contains all of the Innerese sentences (or pointers to them) corresponding to the events in the character's mental-model elaboration graph. This means that retelling the story, with all of the character's inferences, is a simple matter of iterating over the sequence of Innerese sentences and converting them to human-readable English, a standard process in Genesis.

4.2 Answering reading-comprehension questions

At every level of education, from elementary school through college and beyond, instructors ask reading-comprehension questions to determine whether their students understand their reading assignments. The content of the reading and questions may vary drastically—ranging from factual questions about simple fiction stories for elementary school children, to complex thought-provoking questions about technical papers for graduate students—but the concept of reading-comprehension questions is the same.

To assess and demonstrate Genesis's understanding of each character's knowledge in a story, I enabled it to answer DID and WHY questions. DID questions, such

as "Did Carla think that Bobby committed the crime?", demonstrate basic understanding of the story. WHY questions, such as "Why did Carla think that Bobby was guilty?" demonstrate a deeper understanding by engaging the reader to explain characters' motives, as well as to explain their own thought process.

As before, I gave Genesis a basic summary of *Les Misérables*, then asked it DID and WHY questions:

> **Rules:**
>
> If XX commits a crime, then XX is a criminal.
>
> If XX repents, then XX will follow the law.
>
> **Story:**
>
> Inspector Javert patrols the city. Jean Valjean commits a crime. Inspector Javert exits the scene. Jean Valjean repents.

Below are several example questions, with Genesis's answers. (If the grammar in Genesis's answers seems unusual, it is probably because Genesis's built-in Innerese-to-English generator is not perfect; for example, Genesis uses the gender-neutral pronoun "it" if it does not know a character's gender (which tends to be the case with unfamiliar names such as "Jean Valjean").)

> **Does Inspector Javert think that Jean Valjean committed a crime?**
>
> Yes, Inspector Javert thinks that Jean Valjean commits a crime.
>
> **Does Alice think that Jean Valjean committed a crime?**
>
> No, Alice is not a character in this story.
>
> **Does Jean Valjean think that Jean Valjean repented?**
>
> Yes, Jean Valjean thinks that Jean Valjean repents.
>
> **Does Inspector Javert think that Jean Valjean repented?**
>
> No, Inspector Javert does not think that Jean Valjean repents.

Does Jean Valjean think that Jean Valjean will follow the law?

Yes, Jean Valjean thinks that Jean Valjean follows the law.

Does Inspector Javert think that Jean Valjean will follow the law?

No, Inspector Javert does not think that Jean Valjean follows the law.

Why does Jean Valjean think that Jean Valjean will follow the law?

Jean Valjean infers that Jean Valjean follows the law because it repents.

Why does Inspector Javert think that Jean Valjean will follow the law?

Invalid question: It is false that Inspector Javert thinks Jean Valjean to follow the law.

Why does Inspector Javert think that Jean Valjean is a criminal?

Inspector Javert infers that Jean Valjean is criminal because it commits a crime.

Genesis was already equipped with a Question Expert for receiving questions and forwarding them to the appropriate modules to answer them. When a DID or WHY question gets sent to the Perspectives Expert, it first checks the question type (DID or WHY), which is flagged by the Question Expert. Rather than attempting to answer every possible DID or WHY question, I primarily focused on answering questions about each character's understanding of the story, although the Perspectives Expert also has the capability to answer basic factual questions such as "Did Jean Valjean commit a crime?"

After identifying a DID question, the Perspectives Expert checks whether the question is a *belief question*, which takes the general form "Did XX think that EE?", where XX is a character's name and EE is an event. For example: "Did Inspector Javert think that Jean Valjean is a criminal?"

Once Genesis detects a belief question, it divides the question into the character's name (e.g. Inspector Javert) and the target event (e.g. Jean Valjean is a criminal). To answer the question, Genesis searches within the character's mental model, which contains all the explicit events that the character observed (such as "Jean Valjean commits a crime"), as well as the events that the character inferred from the story (such as "Jean Valjean is a criminal"). If no mental model exists, then the character must not be a character in the story, so the default answer is NO: "Does Alice think that Jean Valjean committed a crime?" "No, Alice is not a character in this story." If the character has a mental model, but the target event is not in the mental model, then the answer is assumed to be NO by default because the character did not explicitly observe or infer the target event. Finally, if the target event *is* in the character's mental model—either observed or inferred—then the answer is YES.

The process for answering a WHY question is similar. A WHY question takes the form "Why [belief question]?", such as "Why does Inspector Javert think that Jean Valjean is a criminal?" First, Genesis checks that the answer to the belief question is YES, because it would be invalid to ask "Why does XX believe EE" if character XX does not believe that event EE occurred. If the answer to the belief question is YES, then there are two possibilities: Either the answer is YES because the character observed the event, or the answer is YES because the character inferred the event from something they observed. In the first case, when the character observed the event, Genesis reports an answer of the form "Because XX observed that EE": "Why does Inspector Javert think that Jean Valjean committed a crime?" "Because Inspector Javert observed that Jean Valjean commits a crime." In the second case, when the character inferred the event, the Perspectives Expert uses a method from Genesis's consciousness module to trace back through the elaboration graph to find the explicit story event(s) that resulted in the inference. Then, the answer can take the form "XX inferred that EE, because XX observed that CC", where CC is the cause. For example: "Inspector Javert infers that Jean Valjean is criminal because [Inspector Javert observed that] Jean Valjean commits a crime."

This method differs slightly from how a human reader would answer a question, in

that Genesis has access to complete mental models of everything that every character knows, whereas a human reader might not store as much information while reading a story. However, similar to a human, when asked a WHY question, Genesis first checks the validity of the underlying DID question, then performs a standard backward-chaining search to seek an explanation.

4.3 Conflict detection and explanation: Steps toward conflict reconciliation

Sara Cobb, an expert on narrative-based conflict resolution, notes that in order for two parties to resolve a conflict and reach a reconciliation, each side must first understand and accept the legitimacy of the other's perspective [Cobb, 2013]. One of the first steps in this reconciliation process is to identify, understand, and explain each side's perspective and why they disagree.

For example, when two countries are at war with one another and wish to reconcile, each needs to first understand the other's perspective and accept it as valid, even if they do not necessarily agree. The same can be said of two people in a workplace conflict, or two children arguing on the schoolyard.

At a basic level, many conflicts arise because two people (or countries, or other entities) have opposing beliefs. Thus, the first steps are to identify a conflict and to demonstrate an understanding of both sides. Even if reconciliation is not desired, it can still be valuable to understand situations in which two characters have conflicting views of their world. In the following demo version of *Les Misérables*, the central conflict is that Jean Valjean believes that he is not a criminal, whereas Inspector Javert believes that Jean Valjean is a criminal. I also added a sidekick character who makes the same observations as Inspector Javert, to demonstrate Genesis's behavior when characters have no conflict to resolve.

Rules:

If XX commits a crime, then XX is criminal.

If XX repents, then XX is not criminal.

Story:

Inspector Javert and Officer Pachino enter the scene. Jean Valjean commits a crime. Inspector Javert and Officer Pachino exit the scene. Then, Jean Valjean repents.

When asked the question "Why does Jean Valjean disagree with Inspector Javert?", Genesis reports:

Conflict detected: Inspector Javert and Jean Valjean disagree about Jean Valjean is criminal.

Question: Why did Jean Valjean think that Jean Valjean isn't criminal.?

Jean Valjean infers that Jean Valjean isn't criminal because it repents.

Question: Why did Inspector Javert think that Jean Valjean is criminal.?

Inspector Javert infers that Jean Valjean is criminal because it commits a crime.

Grammar and pronouns aside, Genesis is able to identify a source of conflict between the two characters, and explain why each character believes they are correct. When asked about two characters who do not disagree—"Why does Inspector Javert disagree with Officer Pachino?"—Genesis correctly notes that there is no conflict:

No conflict detected between Inspector Javert and Officer Pachino

From my previous work on answering reading-comprehension questions, the Perspectives Expert already had the ability to detect and sort WHY questions. Now, when it detects a WHY question of the form "Why does XX disagree with YY?", it invokes a conflict explainer that I implemented, which starts by searching for a conflict between the two characters mentioned. For this proof-of-concept system, a conflict is defined as a situation in which two characters have directly opposing views. The conflict could be due to a difference in assumptions or to a difference in what

the two characters observed, but the result is the same: One character believes EE, and the other character believes the negation of EE.

If no conflict is found, Genesis reports "No conflict detected between XX and YY", where XX and YY are the two characters' names. If the Perspectives Expert detects a conflict, it uses its ability to answer WHY questions to explain the conflict. As demonstrated in the example above, if XX thinks EE and YY thinks NOT(EE), Genesis will ask "Why does XX think that EE?" and "Why does YY think that NOT(EE)?", thereby explaining why the two characters disagree. By detecting a conflict and explaining the source of the disagreement, Genesis can take the first steps toward actually resolving the conflict and aiding in reconciliation.

Chapter 5

Contributions

In this thesis, I:

- Developed the Perspectives Expert, enabling the Genesis story-understanding system to **form mental models** of each character's **knowledge of events** in a story by determining which events each character **observes**. Using methods based on **human reasoning**, Genesis can keep track of each character's **location** and **level of consciousness**, also taking into account each character's **common-sense assumptions** and **rules about the world**.

- Employed Genesis's new mental-modeling capability to **retell stories** from different characters' perspectives, **answer reading-comprehension questions** to demonstrate understanding, and perform the first steps of **conflict reconciliation** by detecting and explaining opposing points of view. These applications pave the way for more sophisticated projects such as AI programs that can explain themselves, or a system that resolves conflicts between people, organizations, and nations.

- Highlighted the complexities of determining what each character knows—such as dealing with **one-way information** and **limited senses**—thereby contributing new insights into our computational model of how humans understand stories.

Bibliography

[Call and Tomasello, 1999] Call, J. and Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child development*, 70(2):381–395.

[Cobb, 2013] Cobb, S. B. (2013). *Speaking of Violence: The Politics and Poetics of Narrative in Conflict Resolution*. Oxford University Press.

[Katz, 1997] Katz, B. (1997). Annotating the world wide web using natural language. *Proceedings of the 5th RIAO Conference on Computer Assisted Information Searching on the Internet*, pages 136–159.

[Minsky, 1979] Minsky, M. (1979). K-lines: A theory of memory. *MIT Artificial Intelligence Laboratory Memos*, No. 516.

[Saxe and Kanwisher, 2003] Saxe, R. and Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in theory of mind. *Neuroimage*, 19(4):1835–1842.

[Winston, 2011] Winston, P. H. (2011). The strong story hypothesis and the directed perception hypothesis. In Langley, P., editor, *Technical Report FS-11-01, Papers from the AAAI Fall Symposium*, pages 345–352, Menlo Park, CA. AAAI Press.

[Winston, 2012a] Winston, P. H. (2012a). The next 50 years: a personal view. *Biologically Inspired Cognitive Architectures*, 1:92–99.

[Winston, 2012b] Winston, P. H. (2012b). The right way. *Advances in Cognitive Systems*, 1:23–36.