The Strong Story Hypothesis and the Directed Perception Hypothesis

Patrick Henry Winston
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, MA 02139

Abstract

I ask why humans are smarter than other primates, and I hypothesize that an important part of the answer lies in what I call the *Strong Story Hypothesis*, which holds that story telling and understanding have a central role in human intelligence.

Next, I introduce another hypothesis, the *Directed Perception Hypothesis*, which holds that we derive much of our commonsense, including the commonsense required in story understanding, by deploying our perceptual apparatus on real and imagined events.

Then, after discussing methodology, I describe the representations and methods embodied in the Genesis system, a story-understanding system that analyzes stories ranging from précis of Shakespeare's plots to descriptions of conflicts in cyberspace.

The Genesis system works with short story summaries, provided in English, together with low-level *commonsense rules* and higher-level *reflection patterns*, likewise expressed in English. Using only a small collection of commonsense rules and reflection patterns, Genesis demonstrates several story-understanding capabilities, such as determining that both *Macbeth* and the *2007 Russia-Estonia Cyberwar* involve revenge, even though neither the word *revenge* nor any of its synonyms are mentioned.

Finally, I describe Rao's Visio-Spatial Reasoning System, a system that recognizes activities such as approaching, jumping, and giving, and answers commonsense questions posed by Genesis.

The Right Way

Just about everyone agrees that much has been accomplished since Turing published his seminal paper, *Computer Machinery and Intelligence* (Turing 1950). On the other hand, most would also agree that less has been accomplished than expected. Although applications of Artificial Intelligence are everywhere, we still do not have a computational theory of human intelligence. A team of dedicated first-class engineers can build systems that defeat skilled adults at chess and Jeopardy, but no one can build a system that exhibits the commonsense of a child.

What has been tried? Turing argued that human intelligence is a matter of complex symbolic reasoning. Minsky argues for a multiplicity of ways of thinking coupled into a reasoning hierarchy with instinctive reactions on the lowest level and self-conscious reflection on the highest level (Minsky 2006). Brooks argues that whatever human intelligence is, studying it directly is beyond the state of the art, and we must instead organize systems in layers of competence, starting with the objective of understanding low-level layers that produce insect-level intelligence (Brooks 1991). Still others, in many seminal papers, have suggested that the right way is, for example, through architectural design (Laird, Newell, & Rosenbloom 1987), neural mimicry (McClelland & Rumelhart 1989), or statistical methods (Pearl 1988).

Each of these approaches has made important contributions, especially from an engineering perspective, but none has shown us light at the end of the tunnel, not yet at least.

What is missing, I think, is an approach centered on asking what exactly makes humans different from other primates and from early versions of ourselves. For guidance, I ask about the early history of our species, and I find provocative suggestions in the speculations of paleoanthropologists, especially those of Tattersall (Tattersall 1998).

Basically, Tattersall believes we are symbolic and other primates were not and are not. He says we were not symbolic either, until about 50,000–70,000 years ago. Before that, we were structurally rather modern for perhaps 100,000 years, but during that earlier 100,000 years, like the Neanderthals, all we could do was make simple stone tools and work with fire.

Then, we started making art, and eventually produced the drilled seashell jewelry found in the Blombos Cave, the cave paintings at Lascaux, and the figurines at Brassempouy. Such art, Tattersall believes, requires symbolic thinking and its appearance is evidence of becoming symbolic.

Tattersall argues that we became symbolic rather suddenly, probably in southern Africa, possibly in a population reduced to a few thousand or a few hundred individuals. It was not a matter of slowly growing ability proportional to slowly growing brain size. More likely, it was a evolutionary accident, with nonlinear effects, that unleashed the power of other faculties previously evolved by selection for benefits other than producing human-level intelligence.

Of course, saying we are symbolic does not take us very far toward a computational theory. Chomsky, who fre-

quently cites Tattersall, takes us further by suggesting that we are unique in our ability to combine two concepts to make a third without limit and without disturbing the contributing two (Chomsky 2008). To a linguist, this sounds like the *merge* operation, central to minimalist theories of language.

The Hypotheses

I propose to take Chomsky's hypothesis a step further. I believe that the merge operation gave us the ability to describe events; that we developed the ability to string event descriptions into stories; that we further developed an ability to move backward and forward in remembered stories to explain and predict; that our story processing ability came to include the ability to combine stories into new stories never previously witnessed, from which imagination emerged. Thinking about this kind of thinking has led me to posit the Strong Story Hypothesis:

> **The Strong Story Hypothesis:** The mechanisms that enable humans to tell, understand, and recombine stories separate human intelligence from that of other primates.

Why are stories so important? Because human education is full of stories, starting in modern times with the fairy tales of childhood, through the lessons of history, literature, and religious texts, and on to the cases studied in law, medicine, business, engineering, and science, complemented by the stories told to us by our parents, siblings, and peers. Even learning to follow a recipe when we learn a skill can be viewed as a special case of story understanding.

The pioneering natural-language work of Roger Schank (Schank 1972) presumed that stories are important. Here, with the Strong Story Hypothesis I hypothesize that story understanding is not just important, but rather that story understanding is the centrally important foundation for all human thinking.

Given that story understanding is centrally important, the next question is: Where does the commonsense knowledge needed to understand a story come from? We humans do not get it from the web or from manually built commonsense databases, and even without a desire to understand what makes us different from other primates, depending on the web or other sources of commonsense data is ill advised, because we know a lot we have never been told nor are likely to be told nor are likely to find written down anywhere.

I believe we generate much of what we know as needed, via the interaction of our symbolic and perceptual systems. Sometimes our symbolic system drives our vision system to engage itself on information in the physical world; sometimes our symbolic system drives our visual system to engage itself on an imagined world.

I believe my point of view is well aligned with the work of Ullman on visual routines (Ullman 1996), which in turn was inspired by many psychophysical studies, all of which suggest that our human vision system is a powerful problem solver, not just an input channel. Accordingly, it is natural to draw a picture and move a problem from our symbol-processing faculties to our visual faculties whenever the problem is easier to solve over on the visual side.

We often do not have to draw a picture, however, because imagination is enough. Consider this simple statement–question example: "John kissed Mary. Did John touch Mary?" Everyone seems to answer the question by deploying visual processes on an imagined kiss. Once that is done once, the action-consequence knowledge can be cached as a rule, but being able to get the commonsense answer through perception means you can answer the question, when asked, even if you have not had any sort of kissing education.

Here is a more complex example from personal experience. As a friend helped me install a table saw, he said, "You should never wear gloves when you use this saw." At first, I was mystified, then it occurred to me that a glove could get caught in the blade. No further explanation was needed because I could imagine what would follow. It did not feel like any sort of formal reasoning. It did not feel like I would have to have the message reinforced before it sank in. It feels like I witnessed a grisly event of a sort no one has ever told me about. I learned from a one-shot surrogate experience; I told myself a story about something I have never witnessed; and I will have the commonsense to never wear gloves when I operate a table saw.

From such examples, I posit the Directed Perception Hypothesis:

> **The Directed Perception Hypothesis:** The mechanisms that enable humans to direct the resources of their perceptual systems to answer questions about real and imagined events account for much of commonsense knowledge.

Thus, I believe our inner language enables not only story manipulation but also the marshalling of our perceptual systems, especially our vision perception system, to solve problems on our behalf and produce symbolically cached commonsense rules.

Finally, I believe the Strong Story Hypothesis and the Directed Perception Hypothesis are inseparable, and one without the other loses much of its appeal. Without connection to perception, story understanding reduces to disconnected symbol manipulation by a system that may appear to be quite intelligent, but depends too exclusively on linguistically supplied knowledge. Without connection to story understanding, an otherwise capable perception system can initiate reflex action, but lacks the ability to chain events together, to move backward and forward in such chains, to explain, and to predict.

Genesis

The Strong Story Hypothesis and the Directed Perception Hypothesis are not the whole story, but I believe they are a sufficiently important part of the story to deserve a great deal of research.

I believe that the research should be conducted by looping through the following steps: identify a competence to be understood; formulate a computational problem; propose a computational solution; develop an exploratory implementation; and crystalize emergent principles. These steps are

reminiscent of the methodological-step recommendations of Marr (Marr 1982). The rest of this section illustrates how the steps have guided my research program on story understanding.

Step 1: Competence

What is the competence to be understood? I take it to be that of analyzing stories, such as the following rendering of the plot from Shakespeare's *Macbeth*:

> **Macbeth:** Macbeth, Macduff, Lady Macbeth, and Duncan are persons. Macbeth is a thane and Macduff is a thane. Lady Macbeth, who is Macbeth's wife, is greedy. Duncan, who is Macduff's friend, is the king, and Macbeth is Duncan's successor. Macbeth defeated a rebel. Witches had visions and talked with Macbeth. The witches made predictions. Duncan became happy because Macbeth defeated the rebel. Duncan rewarded Macbeth because Duncan became happy. Lady Macbeth, who is Macbeth's wife, wants to become the queen. Lady Macbeth persuades Macbeth to want to become the king. Macbeth murders Duncan. Then, Lady Macbeth kills herself. Dunsinane is a castle and Burnham Wood is a forest. Burnham Wood came to Dunsinane. Macduff had unusual birth. Macduff fights with Macbeth and kills him. The predictions came true.

I have used simple plot summaries from Shakespeare as anvils on which to hammer out ideas since my earliest work on analogy (Winston 1980). My students and I still use them in our latest, much advanced work because they are easily understood and because they are rich in universally important factors such as power, emotion, consequence, and ties between people. We have found that the same kind of commonsense rules and reflection patterns that work for Shakespeare also work for international conflict, such as the alleged 2007 Russian cyberattack on Estonia's network infrastructure:

> **Cyberwar:** Estonia and Russia are countries. Computer networks are artifacts. Estonia insulted Russia because Estonia relocated a war memorial. Someone attacked Estonia's computer networks after Estonia insulted Russia. The attack on Estonia's computer networks included the jamming of web sites. The jamming of web sites showed that someone did not respect Estonia. Estonia created a center to study computer security. Estonia believed other states would support the center.

Macbeth and *Cyberwar* are representative in length and sophistication of the two dozen stories on which we have focused our attention. Of course, two dozen is a small number, but remember that our ultimate purpose is to understand human understanding, not to engineer systems that only give the appearance of understanding by processing web-sized story sets in ways that shed little light, if any, on human understanding.

What do we mean by *understanding*. After reading stories such as *Macbeth* and *Cyberwar*, everyone has the competence to answer questions like these, ranging from obvious to thought provoking, none of which have explicit answers in the stories themselves:

> Who ends up dead?
>
> Why did Macduff kill Macbeth?
>
> Do the stories involve revenge?
>
> Which story presents a Pyrrhic victory?
>
> Is there a *Macbeth* role in the Russo-Estonia cyberwar?
>
> Is Russia's alleged attack on Estonia's computer networks an instance of *revenge* or *teaching a lesson*?

Step 2: Computational problems

The first computational question is: What representations make it possible to answer questions posed in story understanding? Certainly, knowledge will have to be represented, for without a representation, there can be no model, and without a model, there can be no understanding or explanation.

We could just use some sort of semantic net as a universal representation covering everything, but we felt it would be instructive to see what kinds of knowledge are needed in story understanding, how much of each kind is needed, and how often each kind of knowledge is put to use. Also, we were guided by the principle that refined description tends to expose regularity and constraint.

Regularity and constraint are important, of course, because a model that supports story understanding must involve commonsense and the ability to reflect on the implications of commonsense. This leads to the second computational problem: how do we represent and exploit commonsense and reflective knowledge.

Step 3: Posited solutions

With a view toward building an exploratory system with the ability to answer questions about stories, my students and I—the Genesis Group— anticipated we would need many representations to deal with many kinds of characteristics, relations, and events to be described.

We started with explicit representations for categories whose importance is self evident: class, because what you are determines what you can do (Vaina & Greenblatt 1979); transition, because human reasoning seems to focus on how change causes change (Borchardt 1994), and trajectory, path, and place, because movement along paths is extraordinarily common in language (Schank 1972; Jackendoff 1985).

Next, as we discovered representational needs in exploratory implementation work, we added representations for dealing with coercion (Talmy 1988), cause, goal, persuasion, belief, mood, possession, job, social relations, and time. Then, we added property and role-frame representations as catch-all portmanteaus.

Example sentences leading to the use of our representations are shown in the following list. We anticipate adding other representations as the need emerges.

- **Class:** A thane is a kind of noble.
- **Job:** Duncan was the king.

- **Transition:** Macbeth became the king.
- **Goal:** Macbeth wanted to become the king.
- **Role frame:** Macbeth murdered Duncan.
- **Cause:** Macbeth murdered Duncan because Macbeth wanted to become the king.
- **Persuasion:** Lady Macbeth persuaded Macbeth to want to become the king.
- **Social relation:** Lady Macbeth was Macbeth's wife.
- **Property:** Lady Macbeth was greedy.
- **Possession:** The witches had visions.
- **Mood:** Macbeth became happy.
- **Time:** Then, Lady Macbeth killed herself.
- **Trajectory, path, and place:** Burnham Wood came to Dunsinane.

Genesis also has a representation for *commonsense if–then rules*, for much of story understanding seems to be routine inference making, as knowing that if someone kills someone else, then the someone else is dead. Such rules connect explicit events in the story text with inferred events to form what we decided to call an *elaboration graph*.

Commonsense rule chaining seems necessary but not sufficient for story analysis, because higher-level reflection seems to require search. *Revenge*, for example, is a harm event leading to a second harm event with the actors reversed, possibly with a long chain of intermediate events. I refer to such descriptions as *reflection patterns*. Genesis deploys them using breadth-first search in the elaboration graph. This type of analysis is very different in detail, but inspired by the pioneering work of Lehnert (Lehnert 1981).

Collectively, all our representations constitute Genesis's *inner language*. The representations in the inner language have come to enable description of just the sorts of concepts that would be important for survival, particularly classification, movement in the physical world, relationships in the social world, and various kinds of causation. Perhaps something like Genesis's inner language may eventually shed light on the inner language with which we humans describe the world.

Step 4: Exploratory implementation

With computational problems specified and posited solutions in hand, we set out to develop the exploratory Genesis system.

As a design principle, we decided that all knowledge provided to Genesis—including stories, if-then rules, and reflection patterns—would be provided in English. We were motivated by our debugging philosophy and by the permanence of English; we knew that were we to start over, at least our knowledge base would be reusable.

Given our English-only decision, we had to choose a means to get from English to descriptions couched in our representation suite. Having tried a popular statistical parser, we eventually choose to use the Start Parser, developed over a 25-year period by Boris Katz and his students (Katz 1997), because the Start Parser produces a semantic net, rather than a parse tree, which made it much easier for us to incorporate the Start Parser into a system that translates from English into descriptions in Genesis's inner language.

We also chose to use WordNet (Fellbaum 1998) as a source of classification information. We sometimes augment WordNet with information in English as in "A thane is a kind of noble."

With our Start Parser-enabled translator, we readily express the needed if–then rules in English. Flexibility illustrating examples follow, exactly as provided to Genesis.

- If X kills Y, then Y becomes dead.
- If X harmed Y and Y is Z's friend, then X harmed Z.
- X wanted to become king because Y persuaded X to want to become king.
- Henry may want to kill James because Henry is angry at James.
- If James becomes dead, then James cannot become unhappy.

As the examples show, rules can be expressed *as if–then* sentences or *because* sentences, with or without regular names, and possibly with the modifiers *may* or *cannot*. *May* marks rules that are used only if an explanation is sought and no other explanation is evident. *Cannot* marks rules that act as censors, shutting off inferences that would otherwise be made. In the example, we do not become unhappy when we are dead, even though killing involves harm and harm otherwise causes the harmed to become unhappy.

Reflection-pattern descriptions are a bit more complicated. Here are two versions of *revenge*.

- Revenge 1: X and Y are entities. X's harming Y leads to Y's harming X.
- Revenge 2: X and Y are entities. X's harming Y leads to Y's wanting to harm X. Y's wanting to harm X leads to Y's harming X.

Which is the right version? That, of course, depends on the thinker, so we are able to model specific thinkers by including more or less sophisticated or more or less biased ways of looking at the world.

Genesis reports data about how much each representation is used in the stories read. For example, in working with three Shakespeare plots (*Macbeth*, *Hamlet*, and *Julius Caesar*) and three conflict descriptions (Russia vs. Estonia, Russia vs. Georgia, and North Korea vs. Japan), we found the distribution of representations shown in Table 1. Of course, because the sample size is small, the result is anecdotal, but useful perhaps as a guide to development. We were a little surprised to see only one explicit trajectory, generated by "Birnham wood came to Dunsinane," which leads us to speculate that, at the story level, few such trajectories appear; and we were surprised to see role frames, generated by sentences such as "Macbeth murdered Duncan," in such abundance, leading us to think about subdividing the category. We were not surprised, however, to see many transitions, because many of the commonsense rules produce transitions to new states, as in *If X kills Y, then Y becomes dead*.

Reflection pattern	Fraction in three Shakespeare plots	Fraction in three Cyberwar summaries
Role frame	33%	30%
Transition	15%	12%
Property	9%	9%
Social relation	10%	7%
Mood	5%	8%
Goal	4%	6%
Possession	2%	7%
Time	5%	1%
Belief	1%	4%
Job	3%	1%
Persuasion	2%	1%
Trajectory	1%	0%

Table 1: Various representations occur with varying frequency in the elaboration graphs of six stories in two domains, Shakespeare and Cyberwar.

Equipped with commonsense rules, Genesis produces the elaboration graph of predictions and explanations shown in figure 1. The white boxes correspond to elements explicit in the text; the gray boxes correspond to commonsense inferences. Note that, according to the connections in the graph, Macduff killed Macbeth because Macbeth angered Macduff. Fortunately, we do not always kill the people who anger us, but in the story, as given, there is no other explanation, so Genesis inserts the connection, believing it to be plausible.

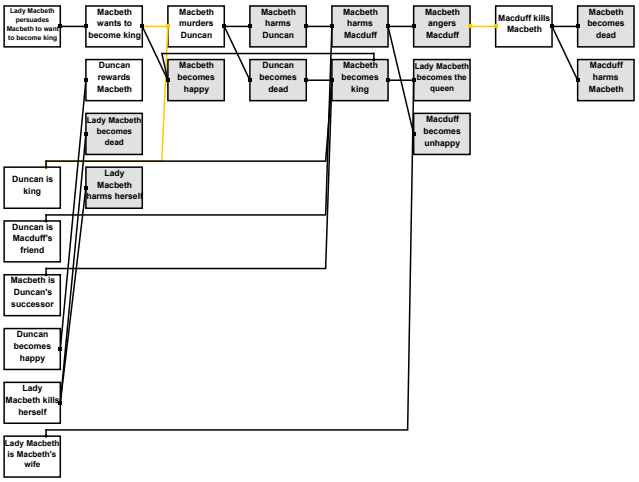

Figure 1: Genesis's story understanding system produces an elaboration graph from commonsense rules together with a story. White boxes indicate information given explicitly in the Macbeth story. Gray boxes indicate information produced by commonsense rules.

Given the elaboration graph, Genesis is ready to look for higher-level concepts of the sort we humans would see in the story but only if we reflect on what we read. Genesis sees, for example, not only *Revenge* but also a *Pyrrhic victory* in the elaboration graph for Macbeth shown in figure 2: Macbeth wants to be king, murders Duncan to become king, which makes Macbeth happy, but then the murder leads to Macbeth's own death.

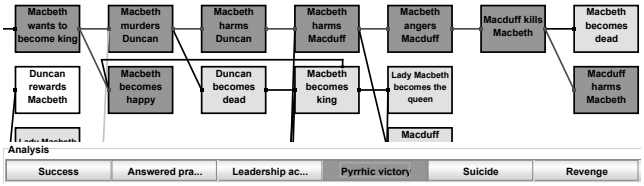

Figure 2: Genesis's story understanding system uses the elaboration graph, together with reflection patterns, to augment the explicit knowledge provided in the story and simple inferences generated using commonsense rules. Here, Genesis discovers a Pyrrhic victory, shown in dark gray.

There is an interesting connection, I think, with the work of Minsky (Minsky 2006). He develops a theory of thinking with six layers: instinctive reactions, learned reactions, deliberative thinking, reflective thinking, self-reflective thinking, and self-conscious reflection. The production of the elaboration graph using commonsense rules seems to me reminiscent of what happens in the bottom three layers. The search for higher-level concepts seems to me reminiscent of what happens in the reflective-thinking layer.

For a more contemporary example, Genesis finds revenge in the elaboration graph produced from a description of the alleged Russian cyberattack on Estonia's network infrastructure, as shown in figure 3.

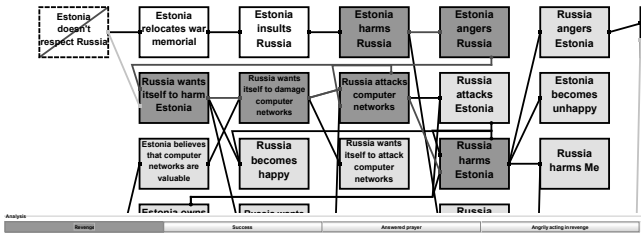

Figure 3: The commonsense rules and reflection patterns honed on Macbeth have broad application. Here, the alleged Russian cyberattack on Estonia reveals an instance of *revenge*, shown in dark gray.

Genesis not only finds revenge, it looks for the acts of harm involved, then uses WordNet to find the most related acts in what the political scientists call the Goldstein index (Goldstein 1992), which enables it to characterize the revenge in *Macbeth* as a tit-for-tat, while the revenge in the Russian cyberattack on Estonia is an escalation.

We have also done some preliminary work on using reflection patterns, such as parallel revenge patterns to help align story elements in preparation for analogical reasoning (Winston 1980; Forbus & Gentner 1989; Gentner & Markman 1997).

To take Genesis to a higher level, we have arranged for the simultaneous reading of stories by two separate persona, which we jocularly call Dr. Jekyll and Mr. Hyde.

Equipped with overlapping but slightly different points of view, Dr. Jekyll and Mr. Hyde see things differently.

In figure 4, for example, Dr. Jekyll concludes that Macduff kills Macbeth in an act of insane violence; Mr. Hyde sees revenge. Both read the same story, but Dr. Jekyll thinks the only reason you would kill someone is that you are insane. Mr. Hyde looks for a reason, and then sees anger. Dr. Jeckll has this rule:

- Henry may want to kill James because Henry is angry at James.

Mr. Hyde has another:

- James may kill Henry because James is not sane.

Social psychologists would say that Dr. Jekyll behaves situationally, more Asian in outlook, because he looks for a situation that has caused a person to do a terrible thing, whereas Mr. Hyde behaves dispositionally, more Western in outlook, because he attributes terrible actions to the characteristics of the actor (Morris & Peng 1994).

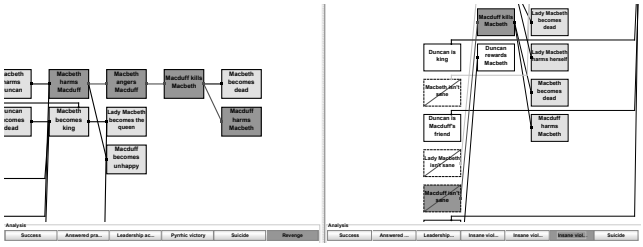

Figure 4: Opinions differ according to culture. One person's act of legitimate revenge is another person's act of insane violence.

Figure 5 shows another example, an example in which Dr. Jekyll sees the alleged Russian cyberattack on Estonia as an act of revenge, because Dr. Jekyll considers Estonia an ally. Mr. Hyde, on the other hand, considers himself an ally of Russia, so the alleged Russian cyberattack is seen as a well-deserved teaching-a-lesson reaction. It is a matter of whose side you are on according to the reflection patterns used in this experiment:

- Revenge: X and Y are entities. X is my ally. X's harming Y leads to Y's harming X.
- Teaching a lesson: X and Y are entities. Y is my ally. X's harming Y leads to Y's harming X.

Step 5: Emergent Principles

At this early stage, it would be a stretch to say principles have emerged. Nevertheless, there have been encouragements and mild surprises.

We were encouraged by the ability of Genesis to work with stories of many types, including not only Shakespeare and conflict in cyberspace, but also simply written fairy tales, law cases, medical cases, and science fiction.

We were surprised that so little knowledge was needed to produce credible performance. Genesis exhibits some characteristics of human story understanding evidenced by

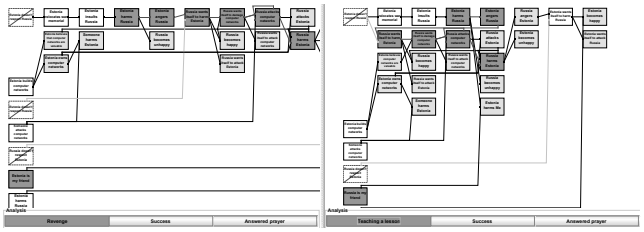

Figure 5: The proper label for the Russia-Estonia cyberattack depends on your point of view. One person's revenge is another person's teaching a lesson.

Genesis's ability to answer a variety of questions about the stories it reads, yet Genesis does its work using only about two dozen commonsense rules and another dozen reflective patterns, several of which, revenge in particular, arose frequently in our experiments, as shown in table 2.

Reflection pattern	Instances in three Shakespeare plots	Instances in three cyberwar summaries
Revenge	4	4
Answered prayer	2	2
Success	1	1
Suicide	4	0
Leadership achieved	1	0
Pyrrhic victory	1	0

Table 2: Some reflection patterns recur frequently; others are infrequent.

What is next

Our work on Genesis has stimulated a great deal of thinking about what to do next, including thoughts about psychological experiments. Here, we report on some areas we think are particularly ripe for technical development, listed from less difficult to more difficult.

Unconstrained use of language

Although we take it as a design constraint that all knowledge shall be in English, our English digestion capability requires us to write simple English. On sufficiently complex sentences, statistical parsers produce defective parses and the Start Parser fails to parse. Moreover, unconstrained English presents problems that lie beyond parsing, such as anaphora resolution.

One way around the simple-English constraint, conceived and developed by Finlayson, lies in what he calls the Story Workbench (Finlayson 2008), a kind of integrated development environment for preprocessing stories. The Story Workbench uses automatic methods together with human annotators who help the automatic methods over the hard parts and supply a measure of interpretation (Finlayson 2008).

Automatic discovery of rules and patterns

At the moment, Genesis's education is direct: we provide all the commonsense rules and all the reflection patterns either directly (in lists) or indirectly (embedded in instructional stories). Of course, humans likewise learn a great deal by being told, but sometimes we form our own commonsense rules and discover and name our own reflection patterns. We want Genesis to do that, too.

On the reflection-pattern level, Finlayson demonstrates how to discover reflection patterns automatically in ensembles of culture-defining stories (Finlayson 2010).

Adding bulldozer computing to understanding

Systems such as IBM's Deep Blue chess player and IBM's Watson Jeopardy contestant demonstrate what can be done by approaching a sufficiently narrow problem with a combination of extremely impressive engineering and massive bulldozer-like computing power. The huge success of such systems has focused a great deal of attention on seeing what can be done by exploiting previously unthinkable quantities of data now readily available on the Internet.

Our work on Genesis has the opposite polarity. We aim to see how little knowledge Genesis needs to reach interesting, humanlike conclusions.

Eventually, we must somehow bring the two approaches together, exploiting large amounts of knowledge with humanlike finesse. This is the hardest problem we face.

Rao's Visio-Spatial Reasoning System

What are the steps involved in addressing the second hypothesis, the Directed Perception Hypothesis? Here the situation is less clear, and I have proportionately less to say. For the moment, even though researchers have worked on image and event understanding for more than half a century, kissing and table-saw operation are too hard to work on now. We can, however, take a big step toward such capabilities by developing systems with the competence to recognize actions such as *jump* and *bounce*, *approach* and *leave*, *throw* and *catch*, *drop* and *pick up*, and *give* and *take*.

To recognize such events, Rao's Visio-Spatial Reasoning System exploits ideas on visual attention (Rao 1998). For each video frame, Rao's Visio-Spatial Reasoning System focuses its attention on the most rapidly moving of the three objects, and computes characteristics of that focal object, such as speed, direction, contact with other objects. In early versions, Rao's Visio-Spatial Reasoning System identified actions, such as the *jump* shown in figure 6, by executing visual routines consisting of manually prepared patterns of attention shift, relative movement, and contact. In later versions, Rao's Visio-Spatial Reasoning System learned the patterns from supervised training examples. Now, efforts are underway to learn action patterns from a combination of unsupervised learning, which identifies common patterns in unlabeled video, and supervised learning, which looks for those common patterns in labeled video (Correa 2011).

We have demonstrated, although not regularly, that we can ask Genesis to ask Rao's Visio-Spatial Reasoning System to answer a question by recalling a video recording from

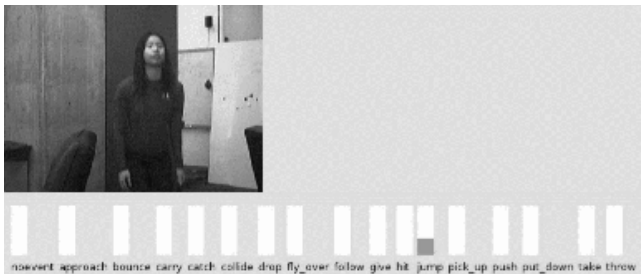

Figure 6: Rao's Visio-Spatial Reasoning System system has learned to recognize *jump* and other actions from a few examples of each. The trained system recognizes that the student is jumping, as indicated by the lighted bar.

memory and reading the answers off of the recalled video using previously learned visual routines.

Suppose, for example, you say that a student gave a ball to another student, and then ask if the other student took the ball. Rao's Visio-Spatial Reasoning System system solves the problem using visual routines that read the answer off a stored, then-recalled scene. Rao's Visio-Spatial Reasoning System recalls the scene shown in figure 7 because, when analyzed visually, the *give* bar lights up. Then, it answers the *take* question by noting that the same scene lights up the *take* bar as well.

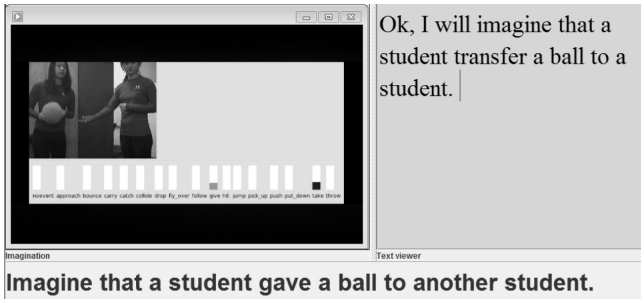

Figure 7: Genesis's language system recalls a situation in which one student gives a ball to another. Because the Rao's Visio-Spatial Reasoning System system sees a *take* in the same sequence, Genesis's language system notes that *give* and *take* co-occur.

Contributions

I first hypothesized that language is important because language enables description, which enables story telling; that story telling is central to education and cultural understanding; and that surrogate experience, in the form of stories, greatly influences culture.

I then hypothesized that language plays a central role in marshalling the resources of our vision system; that language stimulates visual imagination; and that vision is a major problem-solving resource.

Thus, the principal contributions of this paper are the articulation of the Strong Story Hypothesis and the Directed Perception Hypothesis:

- **The Strong Story Hypothesis:** The mechanisms that enable humans to tell, understand, and recombine stories separate human intelligence from that of other primates.
- **The Directed Perception Hypothesis:** The mechanisms that enable humans to direct the resources of their perceptual systems to answer questions about real and imagined events account for much of commonsense knowledge.

These two hypotheses are inseparable. They come together in another hypothesis:

- **The Inner Language Hypothesis:** Human intelligence is enabled by a symbolic inner language faculty whose mechanisms support both story understanding and the querying of perceptual systems.

The Genesis system and Rao's Visio-Spatial Reasoning System are examples of systems built to explore the Inner Language Hypothesis. Those of us who have built the Genesis system and Rao's Visio-Spatial Reasoning System believe we have contributed the following:

- We conceived a research program centered on the Inner Language Hypothesis and the supporting Strong Story Hypothesis and Directed Perception Hypothesis.
- We built a vision system that recognizes human activities such as *approaching*, *jumping*, and *giving*.
- We built a story understanding system with both low-level commonsense and higher-level reflective knowledge, all provided in English.
- We explained how our story understanding system finds concepts such as revenge in stories that never mention the word *revenge* or any of its synonyms.
- We showed how to produce cultural variation in story interpretation through modifications of commonsense and reflective knowledge.

Acknowledgements

This paper was greatly improved by adjustments suggested by Mark Finlayson and anonymous reviewers. The research was supported, in part, by the National Science Foundation (IIS-0413206), the Office of Naval Research (N00014-09-1-0597), the Air Force Office of Scientific Research (A9550-05-1-0321), and the Defense Advanced Research Projects Agency (FA8750-10-1-0076).

References

Borchardt, G. 1994. *Thinking Between the Lines*. MIT Press.

Brooks, R. 1991. Intelligence without representation. *Artificial Intelligence* 47(1-3):139–159.

Chomsky, N. 2008. Some simple evo devo theses: how true might they be for language? evolution of human language: the Morris symposium. Unpublished manuscript.

Correa, T. L. 2011. A model for transition-based visiospatial pattern recognition. Master's thesis, MIT.

Fellbaum, C., ed. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

Finlayson, M. A. 2008. Collecting semantics in the wild: The story workbench, in naturally inspired artificial intelligence. In Beal Jacob, Paul Bello, N. C. M. C., and Winston, P., eds., *Technical Report FS-08-06, Papers from the AAAI Fall Symposium*, 46–53. AAAI Press.

Finlayson, M. A. 2010. Learning narrative morphologies from annotated folktales. In *Proceedings of the 1st International Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts*.

Forbus, K. D., and Gentner, D. 1989. The structure mapping engine: Algorithm and examples. *Artificial Intelligence* 41(1):1–63.

Gentner, D., and Markman, A. B. 1997. Structure mapping in analogy and similarity. *American Psychologist* 52(1):45–56.

Goldstein, J. 1992. A conflict-cooperation scale for WEIS events data. *Journal of Conflict Resolution* 36(2):369–385.

Jackendoff, R. 1985. *Semantics and Cognition*. MIT Press.

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

Laird, J. E.; Newell, A.; and Rosenbloom, P. 1987. SOAR: An architecture for general intelligence. *Artificial Intelligence* 33(1):1–64.

Lehnert, W. 1981. Plot units and narrative summarization. *Cognitive Science* 5(4):293–331.

Marr, D. 1982. *Vision*. Freeman.

McClelland, J. L., and Rumelhart, D. E. 1989. *Explorations in Parallel Distributed Processing*. The MIT Press.

Minsky, M. 2006. *The Emotion Machine*. Simon and Schuster.

Morris, M. W., and Peng, K. 1994. Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology* 67(6):949–971.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: networks of plausible inference*. Morgan Kaufmann.

Rao, S. 1998. *Visual Routines and Attention*. Ph.D. Dissertation, MIT.

Schank, R. C. 1972. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology* 3(4):552–631.

Talmy, L. 1988. Force dynamics in language and cognition. *Cognitive Science* 12(1):49–100.

Tattersall, I. 1998. *Becoming Human*. Harcourt.

Turing, A. M. 1950. Computing machinery and intelligence. *Mind* 59(236):433–460.

Ullman, S. 1996. *High-Level Vision*. MIT Press.

Vaina, L. M., and Greenblatt, R. 1979. The use of thread memory in amnesic aphasia and concept learning. Working Paper 195, MIT Artificial Intelligence Laboratory.

Winston, P. H. 1980. Learning and reasoning by analogy. *Communications of the ACM* 23(12):689–703.