Advances in Cognitive Systems 1 (2012) ?–? Submitted 1/12;; published 1/?

The Right Way

Patrick Henry Winston PHW@MIT.EDU
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, MA 02139, USA

Abstract

I ask why humans are smarter than other primates, and I hypothesize that an important part of the answer lies in the *Inner Language Hypothesis*, a prerequisite to 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. Both the Strong Story Hypothesis and the Directed Perception Hypothesis become more valuable in light of our social nature, an idea captured in the *Social Animal Hypothesis*.

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.

1. Clues from Caves

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 et al., 1987), neural mimicry (McClelland & Rumelhart, 1989), or statistical methods (Pearl, 1988).

© 2012 Cognitive Systems Foundation. All rights reserved.

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.

2. Four Hypotheses

Of course, saying we are symbolic does not take us very far toward a computational theory. Chomsky, who frequently 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, Chomsky's suggestion sounds like the *merge* operation, central to minimalist theories of language. To many practitioners of Artificial Intelligence, Chomsky's suggestion sounds like the capacity to construct complex descriptions, using some sort of language, which leads to the following hypothesis:

> **The Inner Language Hypothesis:** Using a symbolic inner language, we construct symbolic descriptions of situations and events that are far beyond the reach of other primates.

I propose to take the inner language hypothesis a step further. I believe our that we humans 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:** Our inner language enables us to tell, understand, and recombine stories, and those abilities distinguish our 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 and his colleges and students, documented in numerous articles and books (Schank, 1972; Schank & Abelson, 1977; Schank & Riesbeck, 1981; Schank, 1991), presumed that stories are important. So does more recent work on story characterization and story generation, such as found in the numerous papers of authors such as Arthur Graesser (Graesser et al., 2011) and Michael Young (Young, 2007).

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:** Our inner language enables us to direct the resources of our 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.

So far, I have discussed the value of an inner language. What about our outer language? The inner language must have come first, because there is no point in talking if you have nothing to say. Also, there is no point in talking if there is no one to talk to, so it must be important that we are social animals.

> **The Social Animal Hypothesis:** Our social nature amplifies the value of story understanding and directed perception.

Once we somehow acquired an inner language, developing an outer language added substantial benefit and perhaps the inner and outer language coevolved rapidly. Being a social animal does not mean, however, that a sophisticated outer language will emerge, and along with it, an inner language. None of the other bipedal primate species painted caves, and as far as we know, that includes the Neanderthals.

I believe the four hypotheses are inseparable, and one without the others has its possibility destroyed or its potential limited. With no inner language, there can be no story understanding or directed perception. With no story understanding or directed perception, having an inner language might enable some sort of reasoning, but not much educability. 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. Without being a social animal, there would be less learning and less pull on inner and outer language evolution.

3. Five Steps: The Genesis Example

The four hypotheses may not be the whole story, but I believe they are a sufficiently important part of the story to invite a great deal of research. Moreover, I believe that research should proceed by looping through the following five steps: identify the competence to be understood; formulate computational problems; propose computational solutions; 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, by way of example, illustrates how the steps have guided my research program on story understanding.

3.1 First step: Identify the Competence to be Understood

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*?

3.2 Second step: Formulate 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.

3.3 Third step: Propose Computational 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.

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 Schank's student, Wendy 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.

3.4 Fourth step: Develop an 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.

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.

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.

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. Genesis

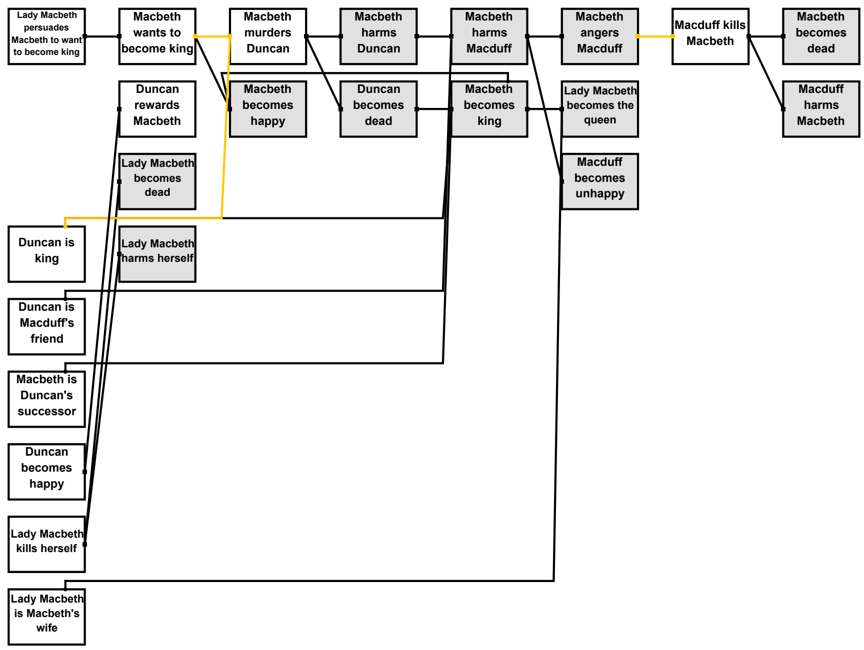

Figure 1. Genesis's story understanding system produces an elaboration graph from commonsense rules together with a story. White boxes indicate information given explicitly in the Macbeth story. Gray boxes indicate information produced by commonsense rules.

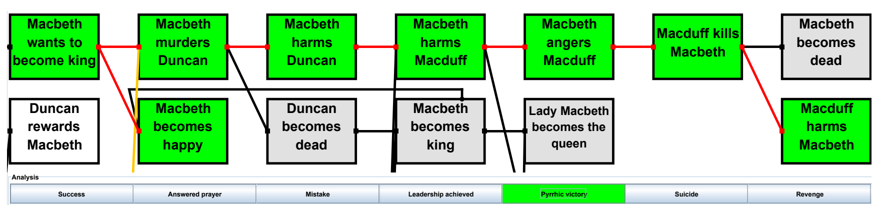

Figure 2. Genesis's story understanding system uses the elaboration graph, together with reflection patterns, to augment the explicit knowledge provided in the story and simple inferences generated using commonsense rules. Here, Genesis discovers a Pyrrhic victory, shown in green.

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.

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. Jekyll has this rule:

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

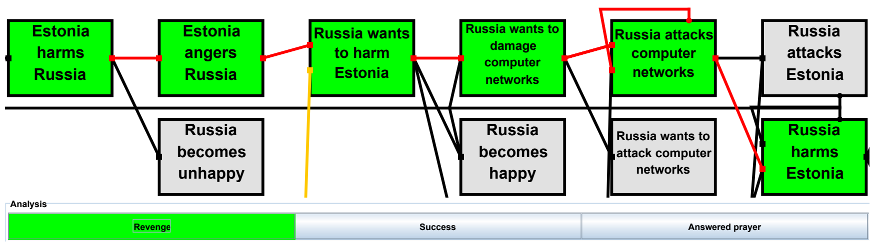

Figure 3. The commonsense rules and reflection patterns honed on Macbeth have broad application. Here, the alleged Russian cyberattack on Estonia reveals an instance of *revenge*, shown in green .

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).

An obvious next step will be to demonstrate culturally guided precedent finding, thus contributing to an important but slowly developing line of research on analogical reasoning. (Winston, 1980; Forbus & Gentner, 1989; Gentner & Markman, 1997; Fay, 2012).

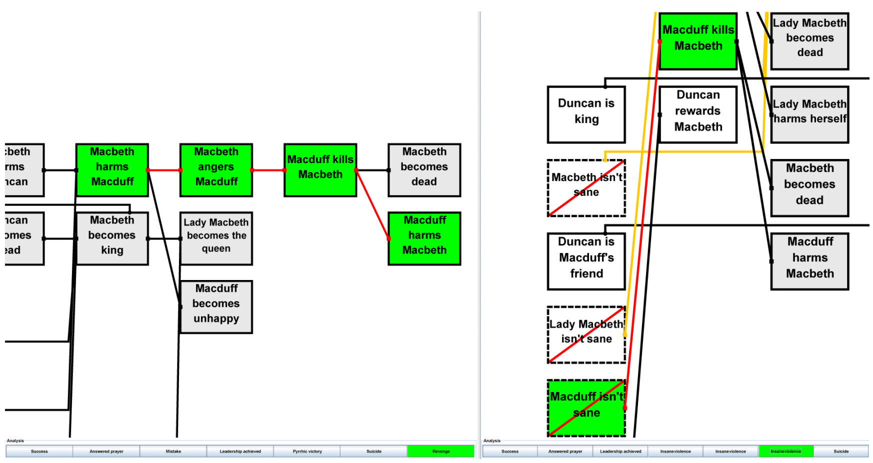

Figure 4. Opinions differ according to culture. One person's act of legitimate revenge is another person's act of insane violence.

3.5 Fifth step: Crystalize 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 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.

3.6 The Way Forward

Good hypotheses lead to better questions. Here I suggest examples, derived from the hypotheses, that are of particular interest to me.

3.6.1 Story understanding

While encouraging, our work on Genesis leaves many hard problem still to be addressed. I list to that are especially important:

How can we build systems that discover their own 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).

How can we add 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 serious engineering, massive computing power, and big data. 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 learn how to handle large amounts of knowledge with humanlike finesse. This is the hardest problem we face.

3.6.2 Directed perception

Vision has received hugely more attention than story understanding, and today's massive computing, big data vision programs can be impressive. Still, even the best of today's people finders occasionally surprise us by misidentifying tree trunks as people, leaving much to be done.

How can we develop vision systems that answer commonsense questions? Fortunately, we seem to be on the leading edge of renewed interest in going beyond object recognition toward visual reasoning. Several groups, with DARPA support, are developing systems that recognize actions such as chase, drop, dig, flee, give, jump, push, and throw. Such systems seem poised to provide play-by-play descriptions of what is happening in front of them, providing a foundation for answering commonsense questions about what particular people are doing and how.

How can story understanding and vision systems learn to communicate? Of course, we can program story understanding and vision systems to communicate, but that soon leads to a deeper question: How do they learn to communicate? In pioneering work, Jeffrey Siskind showed how a basic language system can interact with a simulated vision system (Siskind, 1992); Simon Kirby demonstrated a system in which a syntax emerges from a set of communicating agents (Kirby, 1998); Michael Coen demonstrated a system in which hearing and vocalization subsystems learn an approximation to the mating call of a zebra finch (Coen, 2006); and Jabob Beal offers a mechanism that learns by learning to communicate (Beal, 2007). All these put hands on a very large elephant (a metaphor that demonstrates the amazing power of the human visual reasoning system).

3.6.3 The Inner Language and Social Factors

How can we characterize our inner language? How did our social nature lead to our outer, communication language? Most linguists consider such questions too hard or too underspecified to work on. From my perspective, however, the Genesis system does have an inner language, and that inner language evolves as we ask ourselves what we need to understand stories and drive perception. Accordingly, work on Genesis may well lead to sharpened questions ask about the inner language and the influence of our social nature.

What kind of knowledge is captured in the inner language? If we succeed in developing a story understanding system that reflects understanding approaching that of a human, then its inner language would be an account of the knowledge needed for understanding. Hence, it would be at least a competence theory.

How can the inner language, perception, and the outer language collaborate? Typically, language processors start with word sequences and use statistics to determine how the parsing should be done and how the words should be disambiguated. In situations where perception can play a role, either directed at the real world or an imagined version of the real world, it should be possible to do much better, with the results usefully accumulating in inner-language memories.

4. Contributions

I hypothesized that an inner language is the central element that enables human intelligence. The inner language is important, in part, because it enables description; description enables story telling; story telling is central to education; and surrogate experience, in the form of stories, greatly influences culture.

I then hypothesized that the inner language also plays an important role in marshalling the resources of our vision system; that the inner language stimulates visual imagination; and that vision is a major problem-solving resource. Neither, I argued, would be what they are without the strong influence of our social nature.

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

- **The Inner Language Hypothesis:** Using a symbolic inner language, we construct symbolic descriptions of situations and events that are far beyond the reach of other primates.

- **The Strong Story Hypothesis:** Our inner language enables us to tell, understand, and recombine stories, and those abilities distinguish our intelligence from that of other primates.

- **The Directed Perception Hypothesis:** Our inner language enables us to direct the resources of our perceptual systems to answer questions about real and imagined events account for much of commonsense knowledge.

- **The Social Animal Hypothesis:** Our social nature amplifies the value of story understanding and directed perception.

The Genesis system is an example of a system built to explore the Inner Language Hypothesis. Those of us who have built the Genesis system believe we have contributed the following:

- We conceived a research program centered on the Inner Language Hypothesis and the supporting Strong Story Hypothesis.

- We built a story understanding system with both low-level commonsense and higher-level reflective knowledge, all provided in English.

- We explained how a story understanding system can find 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

Mark Finlayson, Pat Langley, and anonymous reviewers suggested various adjustments that improved this paper. An earlier version was presented at the AAAI Fall Symposium (Winston, 2011). The research describe 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

Beal, J. S. M. (2007). *Learning by learning to communicate*. Doctoral dissertation, MIT.

Borchardt, G. (1994). *Thinking between the lines*. MIT Press.

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 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.

Coen, M. H. (2006). *Multimodal dynamics : self-supervised learning in perceptual and motor systems*. Doctoral dissertation, MIT.

Fay, M. P. (2012). Enabling imagination through story alignment. Master's thesis, MIT.

Fellbaum, C. (Ed.). (1998). *Wordnet: An electronic lexical database*. MIT Press.

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

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

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

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

Graesser, A., Dowell, N., & Moldovan, C. (2011). A computer's understanding of literature. *Scientific Studies of Literature*, *1*, 24–33.

Jackendoff, R. (1985). *Semantics and cognition*. MIT Press.

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* (pp. 136–159).

Kirby, S. (1998). *Language evolution without natural selection: From vocabulary to syntax in a population of learners* (Technical Report). Language Evolution and Computation Research Unit, University of Edinburgh.

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

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

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

McClelland, J. L., & Rumelhart, D. E. (1989). *Explorations in parallel distributed processing*. The MIT Press.

Minsky, M. (2006). *The emotion machine*. Simon and Schuster.

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

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

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

Schank, R. C. (1991). *Tell me a story: A new look at real and artificial intelligence*. Northwestern University Press.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Lawrence Erlbaum Associates.

Schank, R. C., & Riesbeck, C. K. (1981). *Inside computer understanding: Five programs plus miniatures*. Lawrence Erlbaum Associates.

Siskind, J. M. (1992). *Naive physics, event perception, lexical semantics, and language acquisition*. Doctoral dissertation, MIT.

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

Tattersall, I. (1998). *Becoming human*. Harcourt.

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

Ullman, S. (1996). *High-level vision*. MIT Press.

Vaina, L. M., & 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*, 689–703.

Winston, P. H. (2011). The strong story hypothesis and the directed perception hypothesis. *Technical Report FS-11-01, Papers from the AAAI Fall Symposium* (pp. 345–352). AAAI Press.

Young, R. M. (2007). Story and discourse: A bipartite model of narrative generation in virtual worlds. *Interaction Studies*, *8*, 177–208.