

Model-based Story Summary

Patrick Henry Winston

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA, USA
phw@mit.edu

Abstract

A story summarizer benefits greatly from a reader model because a reader model enables the story summarizer to focus on delivering useful knowledge in minimal time with minimal effort. Such a summarizer can, in particular, eliminate disconnected story elements, deliver only story elements connected to conceptual content, focus on particular concepts of interest, such as revenge, and make use of our human tendency to see causal connection in adjacent sentences. Experiments with a summarizer, built on the Genesis story understanding system, demonstrate considerable compression of an 85-element précis of the plot of Shakespeare's *Macbeth*, reducing it, for example, to the 14 elements that make it a concise summary about Pyrrhic victory. Refocusing the summarizer on regicide reduces the element count to 7, or 8% of the original.

1998 ACM Subject Classification I.2.0 General/Cognitive simulation

Keywords and phrases story telling and summarization, story understanding, cognitive modeling

Digital Object Identifier 10.4230/OASICS.CMN.2015.157

1 Vision

Suppose you want a program to summarize a story. How should your program decide what to include and what to leave out? I suggest that people read summaries mainly to acquire useful knowledge in minimal time with minimal effort. Thus, a summary program should focus on knowledge useful as precedent, exclude obvious inferences, but include reflective inferences that help the reader understand how the key elements are connected. Accordingly, a summary program should adhere to several principles reminiscent of the maxims of Grice [5], and in so adhering, a summary program must have an understanding of human story understanding in general and of the summary reader in particular. My students and I have built such an understanding into our Genesis story-understanding system, and we can adjust Genesis to model the knowledge and interests of particular summary readers.

2 Genesis models aspects of story understanding by humans

Much recent work has focused on applications that digest large amounts of data so as to exhibit a kind of intelligence. Google's caption generator [14], for example, is no doubt an engineering marvel, but it sheds little or no light on our human visual faculty. Likewise, IBM's Watson [1] is no doubt intelligent in some ways, but it does not think as we think.

Work on Genesis goes in a different direction. Genesis was developed in the belief that story understanding and telling is the distinguishing feature of human intelligence [15, 16, 17]. The aim in building Genesis is to model aspects of that story understanding and telling feature at the expense of working with story summaries written in simple English of the kind we can get through the START parser [6] and into Genesis's inner language of relations and events.



© Patrick Henry Winston;

licensed under Creative Commons License CC-BY

6th Workshop on Computational Models of Narrative (CMN'15).

Editors: Mark A. Finlayson, Ben Miller, Antonio Lieto, and Remi Ronfard; pp. 157–165

OpenAccess Series in Informatics



OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

One such simple Genesis-readable story is the following précis, which is based loosely on Shakespeare's play, *Macbeth*. It is itself a summary, but it is also an anvil on which to hammer out principles that enable further compression and clarification.

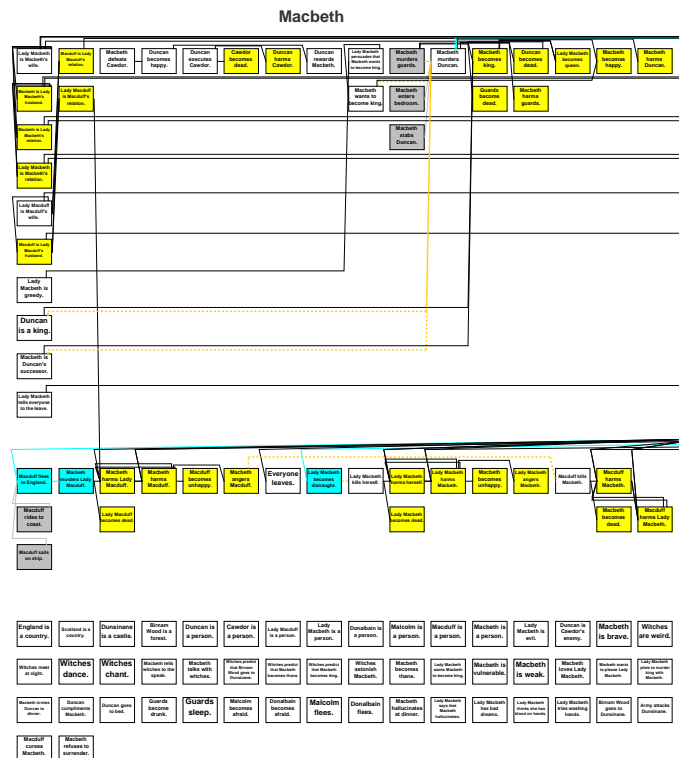
Macbeth précis

Scotland and England are countries. Dunsinane is a castle and Birnam Wood is a forest. Macbeth, Macduff, Malcolm, Donalbain, Lady Macbeth, Lady Macduff, Cawdor, and Duncan are persons. Lady Macbeth is Macbeth's wife. Lady Macduff is Macduff's wife. Lady Macbeth is evil and greedy. Duncan is the king, and Macbeth is Duncan's successor. Duncan is an enemy of Cawdor. Macbeth is brave. Macbeth defeats Cawdor. Duncan becomes happy because Macbeth defeats Cawdor. The witches are weird. The witches meet at night. The witches danced and chanted. Macbeth tells witches to speak. Macbeth talks with the witches. Birnam Wood is a forest. Witches predict that Birnam Wood will go to Dunsinane. The witches predict that Macbeth will become Thane of Cawdor. The witches predict that Macbeth will become king. The witches astonish Macbeth. Duncan executes Cawdor. Macbeth becomes Thane of Cawdor. Duncan rewarded Macbeth because Duncan became happy. Lady Macbeth wants Macbeth to become king. Macbeth is weak and vulnerable. Lady Macbeth persuades Macbeth to want to become the king because Lady Macbeth is greedy. Macbeth loves Lady Macbeth. Macbeth wants to please lady Macbeth. Macbeth wants to become king because Lady Macbeth persuaded Macbeth to want to become the king. Lady Macbeth plots to murder the king with Macbeth. Macbeth invites Duncan to dinner. Duncan compliments Macbeth. Duncan goes to bed. Duncan's guards become drunk and sleep. In order to murder Duncan, Macbeth murders the guards, Macbeth enters the king's bedroom, and Macbeth stabs Duncan. Macbeth becomes king. Malcolm and Donalbain become afraid. Malcolm and Donalbain flee. Macbeth's murdering Duncan leads to Macduff's fleeing to England. In order to flee to England, Macduff rides to the coast and Macduff sails on a ship. Macduff's fleeing to England leads to Macbeth's murdering Lady Macduff. Macbeth hallucinates at a dinner. Lady Macbeth says he hallucinates often. Everyone leaves because Lady Macbeth tells everyone to leave. Macbeth's murdering Duncan leads to Lady Macbeth's becoming distraught. Lady Macbeth has bad dreams. Lady Macbeth thinks she has blood on her hands. Lady Macbeth tries to wash her hands. Lady Macbeth kills herself. Birnam Wood goes to Dunsinane. Macduff's army attacks Dunsinane. Macduff curses Macbeth. Macbeth refuses to surrender. Macduff kills Macbeth.

Given the *Macbeth* précis, Genesis notes and infers several kinds of causal connections. Connections noted are those signaled by the word *because*, the words *leads to*, and the words *in order to* in stories. *Because* signals a direct cause between story elements (Duncan becomes happy because Macbeth defeated Cawdor); *leads to* indicates there is a chain of unstated causes connecting two story elements (Macbeth's murdering Duncan leads to Macduff's fleeing to England); *in order to* explains how something is done (In order to murder Duncan, Macbeth murders the guards, Macbeth enters the king's bedroom, and Macbeth stabs Duncan).

2.1 Genesis deploys various kinds of common-sense rules

In addition to noting explicit causal connections, Genesis produces other causal connections using inference rules, including deduction rules, abduction rules, explanation rules, and



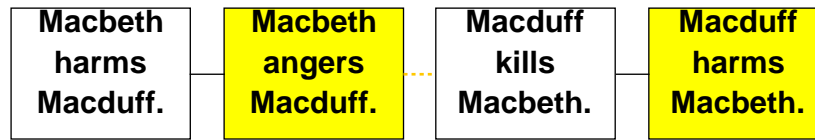
■ **Figure 1** Elaboration graph generated by the *Macbeth* précis. Connections are color coded: deduction rules and explicit *because* connections produce black lines; explicit *leads to* connections produce blue lines; explanation rules produce orange connections. You can expand the diagram if you are using a PDF viewer.

presumption rules. Deduction rules, such as *If x kills y, then y becomes dead*, make connections whenever all their antecedents are in a story. Abduction rules make connections between elements and presumed antecedents. For example, Genesis’s reader model may include the abduction rule *If x kills y, then x must be insane*. Explanation rules make connections only when there is no other known way to explain an element. For example, *Macduff kills Macbeth* is explained by the explanation rule *If x angers y, then y may kill x* and the previously inferred element *Macbeth angers Macduff*. Presumption rules, like abduction rules, make connections between elements and presumed antecedents, but only when there is no other known way to explain an element. Presumption rules, unlike explanation rules, do not require antecedents to be already in place. Abduction rules, explanation rules, and presumption rules are ranked, so that the highest ranking rule dominates in the event multiple rules are available for explaining an unexplained event. We intend to develop a more sophisticated, context-sensitive process.

The noted and inferred causal connections constitute the *elaboration graph* of causally connected elements as shown in Figure 1.

2.2 Genesis discovers concepts by searching for connections

Genesis finds concepts in the elaboration graph by searching for elements that instantiate *concept patterns*. In general, concept patterns include specifications for sequences of causal relations that start and end with particular, specified elements. The concept pattern for



■ **Figure 2** The instantiated revenge concept pattern found in the *Macbeth* précis.

revenge, for example, is just a single such sequence described by *x's harming y leads to y's harming x*. An instantiated revenge pattern is shown in Figure 2.

Remarkably, the elaboration graph, augmented by discovered concept patterns, provides the substrate for developing models of many kinds of story understanding and telling, including question answering, cultural bias in interpretation, instructional telling with a learner model, persuasive telling with a listener model, precedent-based prediction, and as described here, summary.

2.3 We provide common-sense rules and concept patterns in English

My students and I provide Genesis with common-sense rules, concept patterns, and stories; all rules, patterns, and stories are provided in English as indicated in the examples. Our purpose is to establish, by telling, what Genesis needs to know to exhibit a kind of humanlike understanding.

We think it reasonable, at this stage, to tell Genesis what it needs to know. One reason is that much of what we know we learn by being told. Few would have the concept of Pyrrhic victory, for example, without being told. Another reason is that much of what we tell Genesis in experimenting with one story finds use in other stories. Revenge, for example, is revenge not only in *Macbeth*, but also in fairy tales and international conflicts. Yet another reason is that we have done research on learning concept patterns from ensembles of stories [2, 7], and we are engaged in research on learning common sense by mining various textual sources.

3 The Genesis model enables principle-based story summary

Genesis, as a model of story understanding by humans, suggests several principles for summary. Some compress the story provided; others expand the story by adding helpful explanations. All work toward helping the reader to focus on the elements that convey useful knowledge and to grasp how the useful story elements are connected.

In the following, I articulate several such principles, and I explain how those principles are reflected in a model of story summarization by humans. I also show how the Genesis story summarizer, based on that model, performs on a test case.

3.1 The principle of connection

Good precedents exhibit causal connections between events that are likely to be seen again in future situations, thereby enabling understanding, prediction, and control. Accordingly, the Genesis story summarizer preserves those explicit story elements that are involved in causal connections, where the causal connections are either explicit or inferred. Genesis filters out explicit story elements that are neither an antecedent nor a consequent in any kind of causal connection.

The *Macbeth* précis contains 55 sentences, which, when understood by Genesis, expand to 85 explicit story elements, with the expansion caused by separately counting elements that are embedded in compound sentences and explicit causal connections and by adding one to the element count for each explicit causal connection. In what follows, I compare the number of summary elements with the number of explicit story elements for various versions of the Genesis summarizer.

Many of the explicit elements are not involved in causal connections of any kind, explicit or inferred, and thus offer little or nothing by way of constraining precedent. Keeping only those explicit elements that are causal connections and explicit elements that are embedded in Genesis's inferred causal connections produces the following summary in which the START system produces the English, with occasional awkwardness, from Genesis's inner language of relations and events:

Macbeth, with principle of connection

Lady Macbeth is Macbeth's wife. Lady Macduff is Macduff's wife. Duncan is a king. Macbeth is Duncan's successor. Duncan becomes happy because Macbeth defeats Cawdor. Duncan executes Cawdor. Duncan rewards Macbeth because Duncan becomes happy. Lady Macbeth persuades that Macbeth wants to become king because Lady Macbeth is greedy. Macbeth wants to become king because Lady Macbeth persuades that Macbeth wants to become king. In order to murder Duncan, Macbeth murders guards; in order to murder Duncan, he enters bedroom; in order to murder Duncan, he stabs Duncan. Donalbain is Duncan's son. Malcolm is Duncan's son. For Macbeth to murder Duncan leads to Macduff's fleeing to England. In order to flee to England, Macduff rides to coast; in order to flee to it, he sails on ship. For Macduff to flee to England leads to Macbeth's murdering Lady Macduff. Everyone leaves because Lady Macbeth tells everyone to the leave. For Macbeth to murder Duncan leads to Lady Macbeth's becoming distraught. Lady Macbeth kills herself. Macduff kills Macbeth.

Thus, the *principle of connection* allows the Genesis summarizer to reduce the number of summary elements to 34, 40% of the 85 explicit story elements.

3.2 The principle of concept focus

Good precedents tend to be told in a manner that focuses attention on conceptual content because associating a story with its conceptual content is part of what separates novices from domain experts [3, 4]. Accordingly, another version of the Genesis story summarizer includes only explicit elements that lead eventually—via a chain of inferred connections—to an element lying in an instantiated concept pattern.

The elaboration graph plays a central role in this kind of summary because searches in the elaboration graph discover concepts and because searches in the elaboration graph determine which explicit elements are connected to those concepts. Filtering out other elements produces the following *Macbeth* summary:

Macbeth, with principle of concept focus added

The story is about Regicide, Mistake because unhappy, Answered prayer, Revenge, Suicide, Mistake because harmed, Success, and Pyrrhic victory. Lady Macbeth is Macbeth's wife. Lady Macduff is Macduff's wife. Lady Macbeth persuades that Macbeth wants to become king because Lady Macbeth is greedy. Macbeth wants to become king because Lady Macbeth persuades that Macbeth wants to become king.

In order to murder Duncan, Macbeth murders guards; in order to murder Duncan, he enters bedroom; in order to murder Duncan, he stabs Duncan. Macbeth murders Duncan, probably because Macbeth wants to become king, Duncan is a king, and Macbeth is Duncan's successor. For Macbeth to murder Duncan leads to Macduff's fleeing to England. In order to flee to England, Macduff rides to coast; in order to flee to it, he sails on ship. For Macduff to flee to England leads to Macbeth's murdering Lady Macduff. For Macbeth to murder Duncan leads to Lady Macbeth's becoming distraught. Lady Macbeth kills herself, probably because Lady Macbeth becomes distraught. Macbeth becomes unhappy. Macduff kills Macbeth, probably because Macbeth angers Macduff.

Now the summary contains only 30 of the 85 explicit story elements or 35%. Excluded are elements such as *Duncan becomes happy because Macbeth succeeded*, and *Duncan rewarded Macbeth because Duncan becomes happy*. None of the elements involved leads to an element in an instantiated concept.

3.3 The principle of dominant concept focus

Good precedents tend to have a particular purpose and focus attention on one or a few key concepts. Accordingly, yet another version of the Genesis story understander retains an explicit story element only if that element is connected via a chain of inferences to a key concept.

Which of the discovered concepts are the key concepts? There are several reasonable possibilities with which we propose to experiment once we have a large enough corpus of Genesis-readable stories, including concepts that cover a lot of the elements of the story over a long time span, concepts that involve violent acts, such as murder, concepts that excite big emotional reaction, concepts that indicate a dramatic situation, such as those identified by Polti, concepts that the summarizer wants the reader to note, concepts that the summarizer knows the reader wants to note, concepts that are rarely observed, and concepts that involve memorable elements.

For example, in the *Macbeth* précis, Pyrrhic victory dominates all other concepts in the sense that it incorporates the most story elements. Using Pyrrhic victory to summarize, rather than all concepts, Genesis produces the following:

Macbeth, with principle of dominant concept focus added

The story is about Pyrrhic victory. Lady Macbeth is Macbeth's wife. Lady Macduff is Macduff's wife. Lady Macbeth persuades that Macbeth wants to become king because Lady Macbeth is greedy. Macbeth wants to become king because Lady Macbeth persuades that Macbeth wants to become king. In order to murder Duncan, Macbeth murders guards; in order to murder Duncan, he enters bedroom; in order to murder Duncan, he stabs Duncan. Macbeth murders Duncan, probably because Macbeth wants to become king, Duncan is a king, and Macbeth is Duncan's successor. For Macbeth to murder Duncan leads to Macduff's fleeing to England. In order to flee to England, Macduff rides to coast; in order to flee to it, he sails on ship. For Macduff to flee to England leads to Macbeth's murdering Lady Macduff. Macduff kills Macbeth, probably because Macbeth angers Macduff.

The elements that deal with Lady Macbeth's suicide drop out; the number of summary elements is 25, 29% of the explicit story elements.

Memorable elements, incidentally, are readily captured in simple concept patterns that may involve no *leads to* elements, such as this *Memorable event* pattern: *a woman becomes the bishop*. Of course, what constitutes a memorable event may not be so memorable at a different time or place.

3.4 The principle of interpretation transparency

Good summaries do not require readers to guess how the summarizer has reasoned. Accordingly, the Genesis story summarizer is explicit about the assumptions it makes. In particular, the Genesis story summarizer includes not only the consequents of explanation rules, which are explicit in the story, but also the fully instantiated explanation rule, even though the antecedents themselves may be the consequents of deduction rules and not ordinarily included.

For example, the previous two summaries include *Macduff kills Macbeth, probably because Macbeth angers Macduff*. The rationale is that the summarizer, in eagerness to create a more coherent and easily understood story, has added something not completely obvious about how the summarizer has interpreted the story. Thus the summarizer's reasoning is transparent and the reader is relieved of reasoning effort.

3.5 Compression by eliminating details of how actions are performed

Good summaries stick to essentials. Accordingly, the Genesis story summarizer can be directed to eliminate details of how actions are performed, providing further compression.

Impatient readers will not care, for example, about exactly how Macbeth murders Duncan, so the Genesis story summarizer suppresses details about the guards, the bedroom, and stabbing:

Macbeth, with detail suppression added

The story is about Pyrrhic victory. Lady Macbeth is Macbeth's wife. Lady Macduff is Macduff's wife. Lady Macbeth persuades that Macbeth wants to become king because Lady Macbeth is greedy. Macbeth wants to become king because Lady Macbeth persuades that Macbeth wants to become king. Macbeth murders Duncan, probably because Macbeth wants to become king, Duncan is a king, and Macbeth is Duncan's successor. For Macbeth to murder Duncan leads to Macduff's fleeing to England. For Macduff to flee to England leads to Macbeth's murdering Lady Macduff. Macduff kills Macbeth, probably because Macbeth angers Macduff.

With means deleted, the number of summary elements is further reduced to 18, 21% of the explicit story elements.

3.6 Compression using the post hoc ergo propter hoc assumption

Good summaries refrain from making natural inferences explicit because making them explicit is unnatural and annoying. Accordingly, the Genesis story summarizer supposes the reader will instinctively find plausible causal connections between adjacent events.

After this does not mean *because of this* in logic, but we use it nevertheless in telling stories smoothly, dropping explicit cause when proximity makes the cause apparent:

Macbeth, with post hoc ergo propter hoc processing added

The story is about Pyrrhic victory. Lady Macbeth is Macbeth's wife. Lady Macduff is Macduff's wife. Lady Macbeth persuades that Macbeth wants to become king because

Lady Macbeth is greedy. Macbeth wants to become king. Macbeth murders Duncan, probably because Duncan is a king, and Macbeth is Duncan's successor. Macduff flees to England. Macbeth murders Lady Macduff. Macduff kills Macbeth, probably because Macbeth angers Macduff. Macduff.

Processing with post hoc ergo propter hoc transforms *Macduff's fleeing to England leads to Macbeth murders Lady Macduff* to *Macbeth murders Lady Macduff*. With post hoc ergo propter hoc in play, the number of summary elements is 15, 18% of the explicit story elements.

4 Experiments

Using Genesis to summarize Shakespearian play summaries and cyberwar summaries produced the following percentages of summary elements relative to total elements. The Connected column reports the fraction of the explicit story elements that are reported when reporting all and only the elements in the story that are causally connected; the All-methods column reports the fraction of the explicit story elements reported when all of the principles here described are engaged.

	Connected	All methods
Macbeth	40%	18%
Hamlet	41%	14%
Estonia vs. Russia	40%	60%
Georgia vs. Russia	26%	19%

The compression numbers are not dramatic because the test stories are already summaries. The numbers generally drop when limiting the summary to elements that lead eventually to one or more instantiated concept patterns. One exception is *Estonia vs. Russia*. In this summary, one concept pattern is *Aggression of a bully*, a concept pattern that looks for which side the reader is friendly with: *x is my friend. x's angering y leads to y's harming x*. Instantiating that concept pattern brings in *I am Estonia's friend*, a disconnected element, but an element that corresponds to an element in the concept pattern. If the reader happens to be Russia's friend, the concept pattern triggered is *Teaching a lesson* and *I am Russia's friend* is included.

5 Contributions

Work on the Genesis story understanding and telling system has been inspired, in part, by the pioneering work of Roger Shank and his students [8, 9, 10, 11]. Work on Genesis has also been inspired, in part, by paleoanthropologist Ian Tattersall's reflections on what makes us human [12, 13], which led me to the conclusion that story understanding and story telling plays a major role. I have focused here on principles of story summary and shown how those principles are reflected the Genesis story summarizer. In particular, I have:

- Argued that a reader model is a necessary foundation for good story summary
- Identified the principles of connection, concept focus, dominant concept focus, and interpretation transparency.
- Suggested means compression and introduced post hoc ergo propter hoc processing.
- Exhibited an implemented, principle-based summarizer at work on a representative story from the Genesis library, a précis of *Macbeth*, showing a compression of 84%.

References

- 1 David Ferrucci, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John Prager, Nico Schlaefer, and Chris Welty. The AI behind watson—the technical article. *AI Magazine*, 2010.
- 2 Mark A. Finlayson. *Learning Narrative Structure from Annotated Folktales*. PhD thesis, Electrical Engineering and Computer Science Department, MIT, Cambridge, MA, 2012.
- 3 Mark A. Finlayson and Patrick Henry Winston. Intermediate features and informational-level constraint on analogical retrieval. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, pages 666–671, 2005.
- 4 Dedre Gentner and Arthur B Markman. Structure mapping in analogy and similarity. *American Psychologist*, 52(1):45–56, 1997.
- 5 H. Paul Grice. *Studies in the Way of Words*. Harvard University Press, Cambridge, MA, 1989.
- 6 Boris Katz, Gary Borhardt, and Sue Felshin. Syntactic and semantic decomposition strategies for question answering from multiple resources. In *Proceedings of the AAAI 2005 Workshop on Inference for Textual Question Answering*, 2005.
- 7 Caryn Krakauer and Patrick Henry Winston. Story retrieval and comparison using concept patterns. In Mark Alan Finlayson, Pablo Gervas, Deniz Yuret, and Floris Bex, editors, *Proceedings of the 3rd Workshop on Computational Models of Narrative (CMN'12)*, volume 3, pages 119–124. European Language Resources Association (ELRA), 2012.
- 8 Wendy Lehnert. Plot units and narrative summarization. *Cognitive Science*, 5(4):293–331, 1981.
- 9 Roger C. Schank. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, 3(4):552–631, 1972.
- 10 Roger C. Schank and Robert P. Abelson. *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Lawrence Erlbaum Associates, Mahwah, NJ, 1977.
- 11 Roger C. Schank and Christopher K. Riesbeck. *Inside Computer Understanding: Five Programs Plus Miniatures*. Lawrence Erlbaum Associates, Mahwah, NJ, 1981.
- 12 Ian Tattersall. *Becoming Human*. Harcourt, Orlando, FL, 1998.
- 13 Ian Tattersall. Human evolution and cognition. *Theory in Biosciences*, 123(4):371–379, 2010.
- 14 Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. *Cornell University Library digital repository, arXiv:1411.4555 [cs.CV]*, 2014.
- 15 Patrick Henry Winston. The strong story hypothesis and the directed perception hypothesis. In Pat Langley, editor, *Technical Report FS-11-01, Papers from the AAAI Fall Symposium*, pages 345–352, Menlo Park, CA, 2011. AAAI Press.
- 16 Patrick Henry Winston. The next 50 years: a personal view. *Biologically Inspired Cognitive Architectures*, 1:92–99, 2012.
- 17 Patrick Henry Winston. The right way. *Advances in Cognitive Systems*, 1:23–36, 2012.