

# Understanding Stories with Large-Scale Commonsense

**Bryan Williams and Henry Lieberman and Patrick Winston**

Computer Science and Artificial Intelligence Laboratory  
Massachusetts Institute of Technology

## Abstract

Story understanding systems need to be able to perform commonsense reasoning, specifically regarding characters' goals and their associated actions. Some efforts have been made to form large-scale commonsense knowledge bases, but integrating that knowledge into story understanding systems remains a challenge. We have implemented the Aspire system, an application of large-scale commonsense knowledge to story understanding. Aspire extends Genesis, a rule-based story understanding system, with tens of thousands of goal-related assertions from the commonsense semantic network ConceptNet. Aspire uses ConceptNet's knowledge to infer plausible implicit character goals and story causal connections at a scale unprecedented in the space of story understanding. Genesis's rule-based inference enables precise story analysis, while ConceptNet's relatively inexact but widely applicable knowledge provides a significant breadth of coverage difficult to achieve solely using rules. Genesis uses Aspire's inferences to answer questions about stories, and these answers were found to be plausible in a small study. Though we focus on Genesis and ConceptNet, demonstrating the value of supplementing precise reasoning systems with large-scale, scruffy commonsense knowledge is our primary contribution.

## Introduction

Because story understanding is essential to human intelligence, modeling human story understanding is a longstanding goal of artificial intelligence. Here, we use the term "story" to refer to any related sequence of events; a traditional narrative structure is not necessary. Many story understanding systems rely on rules to express and manipulate common sense, such as the rule "If person X harms person Y, person Y may become angry." However, the amount of commonsense knowledge humans possess is vast, and manually expressing significant amounts of common sense using rules is tedious.

Instead, story-understanding systems should be able to make the commonsense connections and inferences on their own, and rule authors should focus on the specifics of a particular domain or story. Commonsense knowledge bases can help achieve this vision. Commonsense knowledge bases express domain-independent and story-independent knowledge, and this knowledge can complement explicit rules by

"filling in the gaps." Our work explores the issues of how and when to use each kind of representation and reasoning.

Genesis is a story-understanding system which models aspects of human story understanding by reading and analyzing stories written in simple English. Genesis can demonstrate its understanding in numerous ways, including question answering, story summarization, and hypothetical reasoning. Prior to this work, Genesis required rule authors to explicitly construct rules that codify all the knowledge necessary for identifying causal connections between story events. For broad-coverage story understanding, explicit construction of all necessary rules is not feasible. ConceptNet (Havasi et al. 2009), a large knowledge base of common sense, helps lessen the burden placed on rule authors. Much of ConceptNet's knowledge has been crowdsourced. ConceptNet includes more than a million assertions, but in this work we draw from 20,000 concepts connected in 225,000 assertions. We have implemented the Aspire system, a new Genesis module which uses ConceptNet's goal-related knowledge to infer implicit explanations for story events, bettering understanding and analysis.

Inferring nontrivial implicit events at a large scale is a new capability within the field of story understanding. Genesis provides a human-readable justification for every ConceptNet-assisted inference it makes, showing the user exactly what pieces of commonsense knowledge it believes are relevant to the story situation. Currently, Genesis is just incorporating ConceptNet's goal-related knowledge, approximately 12,000 assertions in total, into its story processing to demonstrate the viability of this approach. However, we've established a general connection between the two systems so future work can incorporate additional kinds of ConceptNet knowledge.

Despite ConceptNet's admitted imprecision and spotty coverage, we've found it simplifies the process of identifying and applying relevant knowledge at a large scale. However, without judicious use, this imprecision can result in faulty story analyses. Therefore, we developed guiding heuristics which enable Genesis to take advantage of ConceptNet's loosely structured commonsense knowledge in a careful manner. We focus on Genesis and ConceptNet, but we believe the significance of the large-scale application of commonsense knowledge to story understanding we've achieved extends beyond these two systems.

## Background

### ConceptNet

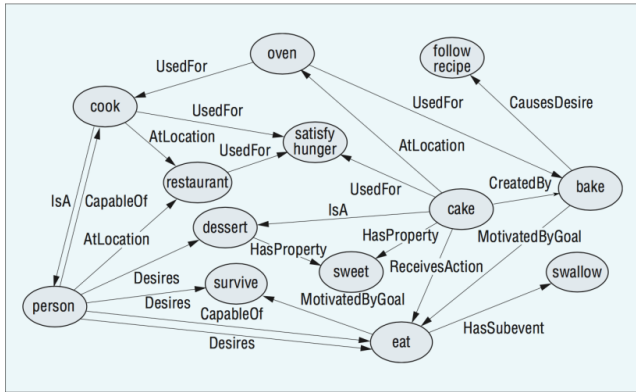


Figure 1: A visualization of a very small part of ConceptNet’s knowledge. Concepts are shown in circles and relations are shown as directed arrows between concepts. © 2009 IEEE. Reprinted, with permission, from “Digital Intuition: Applying Common Sense Using Dimensionality Reduction,” by C. Havasi, R. Speer, J. Pustejovsky, & H. Lieberman, 2009, IEEE Intelligent Systems, 24, p. 4.

ConceptNet is a large semantic network of common sense developed by the Open Mind Common Sense (OMCS) project at MIT (Havasi et al. 2009). There are several versions of ConceptNet, but we used ConceptNet 4 in our work. The latest version of ConceptNet as of this writing is ConceptNet 5, publicly available at <http://conceptnet.io/>. ConceptNet 5 is a larger collection that differs from version 4 primarily by containing “factual” knowledge from WikiDB and other sources in addition to the “pure commonsense”—statements like “water is wet”—expressed in previous versions. In this paper, we use “ConceptNet 4” when discussing attributes of the system specific to that version, and “ConceptNet” at all other times.

ConceptNet represents its knowledge using concepts and relations between them. Concepts can be noun or verb phrases such as “computer”, “breathe fresh air” or “beanbag chair.” There are 27 relations in ConceptNet 4, including “Is A,” “Used For,” “Causes,” “Desires,” “At Location,” and a catchall relation “Has Property.” ConceptNet expresses its knowledge using assertions, each of which consists of a left concept, a relation, and a right concept. For instance, the ConceptNet assertion “computer At Location office” represents the fact that computers are commonly found in offices. Every assertion is associated with a score which represents the confidence in the truth of that assertion. Knowledge reinforced by more sources is assigned a higher confidence score. The knowledge in ConceptNet 4 was mostly crowd-sourced, collected from natural language statements and on-line games-with-a-purpose designed to improve the knowledge base.

ConceptNet assertions are not designed to be interpreted as first-order logic expressions. For instance, ConceptNet

contains the assertion “sport At Location field.” It’s not always true that sports are found on a field, as plenty of sports are played indoors. However, this assertion is still a useful bit of common sense. ConceptNet also does not use strict logical inference. Logical reasoning would combine the reasonable assertions “volleyball Is A sport” and “sport At Location field” to form the conclusion “volleyball At Location field,” which is rarely true because volleyball is usually played indoors or on a beach.

ConceptNet’s knowledge is also not purely statistical. The score for “sport At Location field” is not determined by dividing the number of sports played on a field by the number of sports, as such an approach clumsily ignores context dependency. ConceptNet does apply some statistical machinery to form new conclusions from large numbers of assertions, but it uses human-supplied generalizations as input rather than pointwise data like occurrences of words in documents.

ConceptNet is intentionally “scruffy” in nature (Minsky 1991), embracing the chaotic nature of common sense. Assertions can be ambiguous, contradictory, or context-dependent, but the same is true of the commonsense knowledge humans possess, albeit to a different degree. ConceptNet’s lack of precision does mean applications that use it must be careful in exactly how they apply its knowledge. We discuss our heuristic approach in “The Aspire System” section of this paper.

### Genesis

Genesis is a story-understanding system started by Patrick Winston in 2008, and is implemented in Java. Stories are input to Genesis via simplified natural language sentences and parsed into story events by Katz’s START parser (1997). We use the term story event to refer to story sentences, phrases, or inferences that Genesis extracts as a unit of computation.

Genesis can summarize stories, answer questions about them, perform hypothetical reasoning, align stories to identify analogous events, and analyze a story from the perspective of one of the characters, among many other capabilities (Winston 2014; Holmes and Winston 2016; Noss 2016). To perform intelligent analyses of a story, Genesis relies heavily on making inferences and forming causal connections between the story’s events. If these connections or inferences are of poor quality, the analyses suffer. Prior to this work, commonsense knowledge was obtained exclusively through rules, putting a large burden on the rule author. The same ruleset is likely useful for many stories, especially within the same genre, so the author may be able to reuse or modify previously constructed rulesets. Still, separating general, often banal commonsense knowledge from Genesis rules frees the rule author to concentrate on domain specifics and high-level story patterns. ConceptNet is considerably larger than any previously constructed Genesis rule set.

### The Aspire System

Goals are essential to human nature. Humans are constantly forming goals and taking actions towards them, and Schank long ago recognized the importance of identifying character

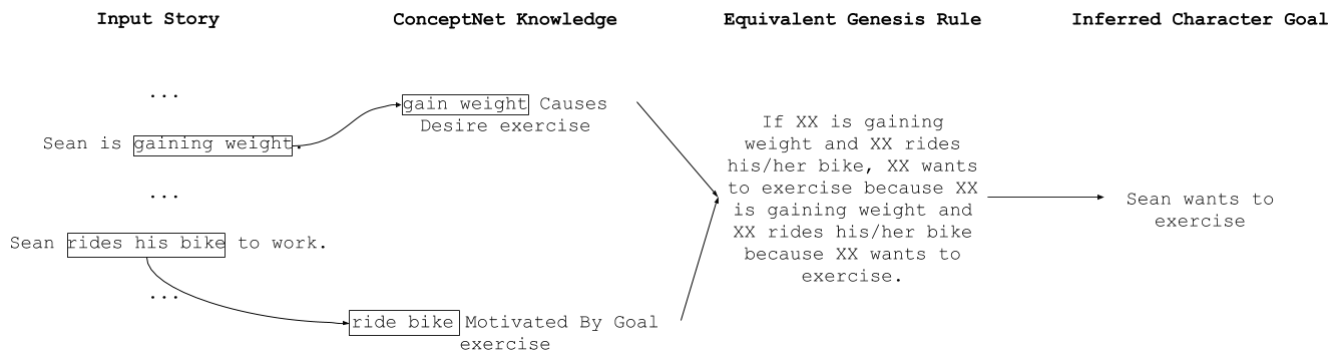


Figure 2: The inference and equivalent rule formulation Aspire makes from the example explicit story events “Sean is gaining weight” and “Sean rides his bike to work.” The inferred character goal “Sean wants to exercise” connects these two story events.

goals when processing natural language (1977). Low-level goals can change the way the brain processes information, altering what people pay attention to and what people remember (Kenrick et al. 2010). At a more conscious level, goals are an effective motivational tool (Locke and Latham 2006). Because goals drive humans at both a subconscious and conscious level, goals are prevalent in the stories humans tell as well. Enabling Genesis to better understand goals, their causes, and actions taken towards completing them improves its comprehension of goal-oriented stories.

ConceptNet is a great resource for analyzing goals because it contains a multitude of commonsense knowledge about both what causes and fulfills goals. The relations “Causes Desire,” “Motivated By Goal,” “Used For,” and “Has Subevent” are all particularly relevant to the goal domain. We have given Genesis the means to leverage this knowledge while processing a story to perform goal-related inference.

We introduced the Aspire system, a Genesis module which analyzes characters’ goals, their causes, and the actions taken to complete them. Aspire uses approximately 12,000 goal-related assertions from ConceptNet, but can operate using explicitly authored rules as well. With Aspire and ConceptNet, Genesis can arrive at a more complete understanding of a story with less effort from the rule author.

## Implementation

Aspire works by maintaining a set of candidate character goals as Genesis sequentially reads the story input, adding to the set when it detects that a story event might cause a character in the story to have a goal. Aspire analyzes each story event to see if it causes a candidate character goal. It also checks if each story event is a goal contribution, an action taken by a character to accomplish a previously identified candidate character goal. Candidate character goals are kept merely as candidates and do not affect other Genesis modules until Aspire sees a goal contribution for a candidate character goal, at which point the goal is “confirmed” and the inference is inserted into the story. Aspire also analyzes its own inferences for goal causation and contribution, allowing its inferences to build off one another. Note that

we use the terms “candidate character goal” and “candidate goal” interchangeably.

As an example, consider a story about a man named Sean that contains the event “Sean is gaining weight.” In this scenario, Aspire checks if this event causes a goal by examining the rules Genesis is given and by consulting ConceptNet. ConceptNet knows gaining weight causes the desire of exercising, so Aspire receives the assertion “gain weight Causes Desire exercise” from ConceptNet, among other relevant assertions. Aspire adds the corresponding candidate character goals to its set, including “Sean wants to exercise,” and continues reading the story.

In addition to goal causation analysis, Aspire also performs goal contribution analysis on every received story event. During goal contribution analysis, Aspire checks if the current story event contributes to any of the candidate goals in its set. To detect a contribution between an event and candidate goal, Aspire first tries to match the event to the goal using traditional Genesis matching, which does not involve ConceptNet. If this match succeeds, Aspire concludes that the event contributes to the candidate character goal. If matching fails, ConceptNet is consulted.

Suppose the story contains some subsequent event “Sean rides his bike to work.” During goal contribution analysis of this event, Genesis matching between bike riding and Sean’s candidate goal of exercising fails. Therefore, Aspire extracts the concept “ride bike” from the event and “exercise” from the candidate character goal. Aspire then queries ConceptNet to see if any of “ride bike Motivated By Goal exercise,” “ride bike Used For exercise,” or “exercise Has Subevent ride bike” are true. If ConceptNet confirms any of these assertions, Aspire concludes the event contributes to the candidate character goal; otherwise, with both Genesis matching and ConceptNet failing to form a connection between the event and the goal, Aspire concludes the event does not contribute to the goal. In this goal contribution analysis, ConceptNet confirms “ride bike Motivated By Goal exercise,” and Aspire links Sean’s bike riding to his goal of exercising. While there is no guarantee that these two events are causally related, storytelling convention and commonsense reasoning indicate it’s likely they are. Aspire is designed to

generate plausible inferences, not indisputable ones.

When Aspire concludes that a story event contributes to a candidate character goal, Genesis forms the associated causal connections and adds new inferred events to the story. Consider once again the Sean example. Having analyzed “Sean is gaining weight” and “Sean rides his bike to work,” Genesis adds “Sean wants to exercise” to the story. Genesis also forms causal connections between the inferred goal’s cause, the inferred goal, and the goal contribution action. The completion of the goal (“Sean exercises”) is instantiated and added to the story as well. Genesis explicitly adds the inferred goal and inferred goal completion to the story so that all modules, including Aspire, can process these inferences and draw additional conclusions. For example, Sean exercising could contribute to a candidate character goal “Sean wants to lose weight” which was formed earlier in the story. In this way, Aspire’s inferences recursively enable additional Aspire inferences.

Importantly, the character goal and character goal completion are only inserted into the story once Aspire sees an action taken that contributes to the character goal. Candidate character goals on their own do not affect the story in any way—a character action must first prove a candidate goal credible before Genesis acts on Aspire’s analysis. If Aspire never sees such an action, the candidate goal remains a potential interpretation of the story’s events that lacks sufficient evidence. Aspire operates in this way because events that can cause a goal do not always cause that particular goal. In the Sean example, weight gain does not necessitate a desire to exercise. It’s possible that Sean’s weight gain caused him to diet instead of exercise, or maybe he didn’t take any response at all. Aspire’s defensive manner also helps prevent misapplication of ConceptNet’s knowledge. ConceptNet’s inexact nature allows its knowledge to be widely applicable, but its knowledge can also be ambiguous, context-dependent, or otherwise imprecise. Therefore, Aspire requires seeing both a goal’s cause and a goal’s effect before inferring the goal to help ensure it has formed a correct inference.

Extracting relevant ConceptNet concepts from a story event is not always straightforward. For example, in the sentence “Matt buys a book,” it would be appropriate for Aspire to extract both “buy book” and “book” as concepts to see what goals these may cause or complete. However, if the event were instead “Matt loses his book,” while it would still be appropriate to extract “lose book” as a concept, “book” by itself would not be appropriate because ConceptNet provides knowledge about possessing a book, not losing it.

As this example shows, verbs and verb phrases are simpler to extract than nouns and noun phrases because their meaning is less dependent on the rest of the sentence. Therefore, nearly all of the concepts Aspire extracts from story events are verbs or verb phrases rather than nouns or noun phrases. We developed heuristics which specify the behavior for concept extraction and matching based on the part of speech of the concept, the transitivity of the verb in a verb phrase concept, and whether a noun concept is a proper noun. These heuristics were devised by examining patterns in ConceptNet data. More specifically:

- Nouns or noun phrases are only extracted if the verb is a form of “be,” “have,” or “feel” (e.g. “compassion” is extracted from “Matt has compassion”).
- Proper nouns never appear in extracted verb phrase concepts, as ConceptNet generally does not contain knowledge about specific people, places, or things.
- When a verb takes an object that isn’t a proper noun, verb phrases are extracted rather than sole verbs because the knowledge about the sole verb may not be relevant when the object is taken into account (e.g. “fight” vs. “fight inflation”).
- Modifying adverbs can be included in extracted verb phrase concepts, allowing Aspire to query ConceptNet for “work hard” in addition to “work” upon reading “Matt works hard.”
- The transitivity of the verbs in the goal causation and goal contribution need not match. Aspire can detect that “Matt sings” contributes to “Matt wants to make music.”
- Aspire assumes that, in an assertion, the subject and optional object the left concept takes are consistent with the subject and optional object the right concept takes. For example, if Aspire reads that “Matt loves Helen” and learns from ConceptNet that “love Causes Desire kiss,” Aspire adds “Matt wants to kiss Helen” as a candidate goal. It does not hypothesize that Helen wants to kiss Matt, Matt wants to kiss someone else, or someone else wants to kiss.

While these heuristics are not infallible, they help ensure Aspire is using data appropriate to the story in the correct manner. We describe the extraction and matching algorithms in more detail in our earlier work (Williams 2017).

## An Example Analysis

In this section we describe the performance of the Aspire system on an example story. Genesis justifies its conclusions with human-readable text that describes the relevant knowledge obtained from ConceptNet. We also give an example of how Aspire’s capabilities allow other Genesis features to perform better.

The following is a simple retelling of the Mexican-American war, a conflict which occurred in the 1840s. Because of the limitations of the START parser, stories are input to Genesis using simple English.

The United States is a country. Mexico is a country. The year is 1846. Manifest Destiny is popular in the United States. The United States has ambition, has greed, and wants to gain land. Mexico and the United States disagree over borders. The United States moves into the disputed territory. Mexican forces attack the United States. The United States declares war on Mexico. Winfield Scott leads the United States army. The United States battles Mexico. The United States triumphs by capturing Mexico City. The United States defeats Mexico and wins the war. The Treaty of Guadalupe Hidalgo officially ends the war.

This summary of the war is plain, brief, and high-level, resembling an elementary school textbook chapter summary.

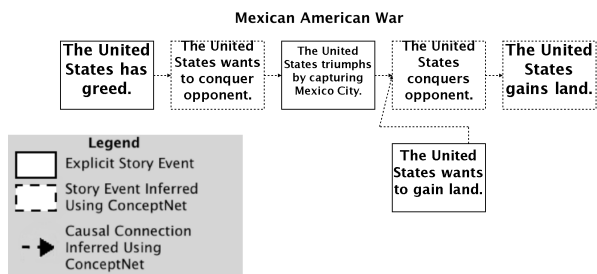


Figure 3: A subset of the elaboration graph produced by Genesis for the Mexican-American war story. All events and connections inferred with ConceptNet knowledge are shown using dotted lines. Genesis was not given any author-defined rules to help analyze this story, instead relying entirely on ConceptNet.

This example focuses on ConceptNet, exploring how Genesis analyzes this story using Aspire when it is not given any rules from an author. There is one goal explicitly stated in the story—“The United States wants to gain land”—but Genesis is not told how this relates to any of the other story events.

A subset of the elaboration graph depicting Genesis’s analysis of the Mexican-American war story is shown in Figure 3. The elaboration graph is a central Genesis display which shows the inferences and causal connections Genesis has made. Each story event is shown in a box, and arrows between boxes point from cause to effect. Events and connections inferred using ConceptNet knowledge are shown using dotted lines.

Aspire allows Genesis to connect the United States’ ambition and greed to its actions taken against Mexico. The inferences Genesis makes are just one possible interpretation of the story’s events, and the English rendering of some of its inferences is not perfect. Still, the inferences are salient because the effective application of relevant knowledge is the focus of our work and this example. Our design strives for plausible inferences, not certain ones.

ConceptNet helps Genesis connect the United States’ ambition and greed to its desire of conquering its opponent, which it completes when it triumphs by capturing Mexico City. Conquering its opponent allows it to gain land, a goal explicitly stated in the story. The country’s greed makes it want to get something, and this goal is accomplished when the United States gains land, an example of an Aspire inference enabling Genesis to make an additional Aspire inference. There was no mention of conquering anywhere in the story text, but Genesis has inferred that this concept is relevant. Genesis was able to make all of these inferences and connections without any author-defined rules; instead, it’s relying just on ConceptNet knowledge.

Aspire tracks the ConceptNet knowledge it uses while analyzing a story. Genesis uses this data provenance to display justifications for its analysis to the user, increasing transparency by showing exactly why Genesis believes a particular piece of ConceptNet knowledge is relevant to a situation. Some of the justifications for the Mexican-American

Figure 4: Two of the justifications Genesis produces when analyzing the Mexican-American war story. Genesis justifies every ConceptNet-assisted inference with the relevant assertion from ConceptNet, increasing transparency. The confidence score accompanies each assertion.

Figure 5: A series of questions and Genesis answers regarding the Mexican war story. Genesis’s grammar in its English generation is not perfect, but the content is salient.

war story are displayed in Figure 4.

Causal connections are the base of the majority of Genesis’s features because they form its reasoning substrate. The ConceptNet-assisted causal connections and inferences Aspire forms improve the performance of many Genesis features even though the rule author has given Genesis less information. Question answering is an example of such a feature. Two questions about the Mexican-American war story and Genesis’s associated answers are displayed in Figure 5.

When Genesis is asked “How did the United States gain land?”, it references two story events that were never explicitly stated, albeit with imperfect grammar. When Genesis is asked about one of these inferences, it responds with further explanation, citing specific explicit story events. If Aspire were absent, Genesis would be unable to answer “Why” or “How” questions about this story as no rules were authored for it to use.

## Evaluation

Is the Aspire system capable of producing plausible, human-like inferences? We conducted a small study to compare Aspire’s analyses with human reading comprehension. Evaluating Aspire for accuracy on a large corpus of stories in unrestricted natural language was not feasible due to limitations of the START parser and ConceptNet’s limited cov-

erage. Given these terms, rather than test Aspire’s comprehensive story understanding ability, our goal was to evaluate whether participants considered Aspire’s answers to simple reading comprehension questions plausible. The results were promising. Participants’ answers to the questions were roughly compatible with Aspire’s, and they found Aspire’s answers plausible.

Five male participants took part in the experiment. The participants were all in their early 20s, were racially diverse, and had varying academic backgrounds. We presented two example stories, one fiction and one nonfiction. The nonfiction story replaced real-world proper names with fictional names to avoid participants relying on their prior knowledge of historical events rather than what was explicitly mentioned in the text.

First, the participant read each story and answered the questions. Then, they were shown Aspire’s answers and rated them for plausibility on a Likert-5 scale. Each participant was asked five questions about each story. Participants were also asked up to three follow-up questions depending on their responses to the original five. The participants then rated thirteen of Aspire’s responses. Because we were testing Aspire’s application of ConceptNet knowledge, all questions focused on ConceptNet-assisted inferences. We removed any grammatical errors and oddities from Aspire’s responses so that these would not distract the participant.

The results from the evaluation indicate that the participants found Aspire’s inferences largely plausible. The average rating given by participants was 4.77 on a scale of 5, suggesting that participants do tend to agree with Aspire’s answers to the questions.

## Discussion and Future Work

Much of ConceptNet’s knowledge remains untapped, as Aspire only uses the “Causes Desire,” “Motivated By Goal,” “Used For,” and “Has Subevent” relations. There are numerous other relations that would provide valuable information to Genesis, with “Causes” being the most enticing one. Causal knowledge could directly wire into Genesis’s story processing so Genesis could identify general causal relationships in the story, moving beyond the goal-oriented connections Aspire forms.

Aspire operates using an important simplifying assumption which does not always hold: if a goal has had one or more actions taken towards completing it, the goal has been completed. As an example, suppose Aspire has formed the candidate character goal “Matt wants to dance.” Aspire would connect the event “Matt plays music” to this goal because ConceptNet contains the knowledge “play music Motivated By Goal dance.” Because the goal has had an action taken towards it, Aspire would instantiate both the inferred goal “Matt wants to dance” and the inferred goal completion “Matt dances.” Instantiating the goal completion at this point is premature, though, and is incorrect if the story later indicates that Matt did not get to dance after all. We also do not consider more complex configurations where there are multiple interlocking or inhibiting goal structures.

It would be much better if, for any given goal, Aspire could distinguish which sorts of actions complete that goal

from which sorts of actions merely contribute to it. Differentiating goal contribution from goal completion is just one example of the more robust understanding we’d like Aspire to reach. Ideas from Schank’s Conceptual Dependency Theory could be useful in achieving deeper understanding, especially given recent work exploring how to crowdsource the requisite knowledge (Schank 1975; Macbeth and Grandic 2017). Adapting approaches and principles from Segmented Discourse Representation Theory (SDRT) could also prove beneficial (Lascarides and Asher 2008).

Alternative sources of commonsense such as Cyc and Wechilid (Guha and Lenat 1994; Tandon et al. 2014) may go beyond ConceptNet in providing more finely grained knowledge for story understanding. They could also assist in processing polysemy, homonymy, and the like. ConceptNet’s simple design chooses to not capture these language intricacies, although they are addressed by some later versions of ConceptNet (Havasi, Speer, and Pustejovsky 2010; Chen and Liu 2011). Multiple commonsense knowledge bases could even be used in conjunction. We are not bound to Genesis either; other rule-based and logical systems certainly warrant investigation. Our broader focus is not on Genesis, ConceptNet, or Aspire, but instead on supplementing precise reasoning systems with large-scale, scruffy commonsense knowledge. This valuable combination results in both broad coverage and a significant depth of understanding.

## Related Work

While many have applied commonsense knowledge to story understanding, we are not aware of any story understanding system that incorporates common sense at as large a scale as this work. Prior work in this area tends towards the creation of a small amount of handcrafted commonsense rules rather than harnessing a substantial amount of common sense for use in many different contexts. Note that the work described in this section is far from a complete survey; instead, we focus on several closely related efforts and trends.

Gordon (2016) recently framed the problem of forming commonsense interpretations as a process of logical abduction, building on the ideas of Hobbs, Stickel, Martin, and Edwards (1988). He developed a model which reads a small story and chooses the more likely of two possible relevant story inferences. The model consults 136 hand-coded probabilistic commonsense axioms to generate scored hypotheses. The model performed well on the Triangle-COPA benchmark set, but it’s not clear how scalable the approach is. The model requires carefully constructed commonsense axioms, and Gordon does not propose a way to generate these axioms at scale.

Blass and Forbus (2017) have taken a larger scale approach to commonsense causal reasoning through their analogical chaining formalism. Their reasoning system is initialized with a Cyc ontology but can also take natural language instruction. It answers questions from the COPA dataset. While they focus only on selecting plausible consequents of events, not general story understanding, their approach shares similarities with ours.

The field of machine reading, focused on the application of machine learning techniques to natural language understanding, is rapidly developing. Notable topics of interest include semantic role labeling, named-entity recognition, and question answering, the area most closely related to this work (Young et al. 2017). Popular benchmark question answering datasets from Stanford, Facebook, and Google DeepMind (Rajpurkar et al. 2016; Weston et al. 2015; Hermann et al. 2015) are routinely used to evaluate current models. However, the answer for every question in these datasets is always explicitly stated in the input text (besides when the question has a yes or no answer, in which case the story explicitly contains all relevant information needed to answer the question). Little inference and commonsense knowledge is required, as all the questions can be answered by combining information in different sections of the input text. These datasets are a good first step towards more general intelligence, but natural language understanding systems must be able to form inferences using background knowledge. The recently introduced RACE dataset (Lai et al. 2017) requires significantly more commonsense reasoning in its question answering tasks than prior benchmarks, but has yet to gain popularity. The Story Cloze test, in which a system is tasked with reading a four-sentence story and choosing the more plausible ending from two choices, is another popular means of evaluation (Mostafazadeh et al. 2017). Importantly, all of these datasets and evaluation schemes test the ability of a system to select a plausible inference from a set of choices, not generate its own. Aspire’s capacity to generate inferences distinguishes it from popular natural language understanding work, but also complicates evaluation.

## Conclusion

Aspire brings us closer to the goal of applying Genesis to a corpus of naturally occurring text. By consulting ConceptNet, a large commonsense knowledge base composed of human-submitted assertions, Genesis can make many more inferences and causal connections. Rule authors can now focus significantly more on domain-specific rules describing niche information, letting ConceptNet fill in gaps with universal common sense. This technology is quite valuable in any domain that heavily depends on precedent, including law, medicine and business.

We found that ConceptNet’s relaxed data format simplified the process of forming plausible inferences, but also required developing heuristics to help prevent imprecision from resulting in incorrect inferences. Encouraged by these initial results, we look forward to incorporating additional large-scale common sense into Genesis processing. While the union of ConceptNet and Genesis has proven effective, our primary contribution is the general approach taken rather than the specifics. Several other rule-based systems, commonsense knowledge bases, and inference schemes merit exploration as well.

Many AI systems face the dilemma of how to simultaneously achieve breadth and depth. They desire the ability to handle a wide range of subject matter, but also strive to perform analysis at a meaningful and appropriately complex

level. Trying to squeeze all these capabilities into a single representation and a single inference procedure might be a fool’s errand. Instead, we are inspired by Minsky’s recommendation of having multiple representations and multiple inference procedures (1986). Our work explores how to get them to work together, letting each do what they do best.

In the case of story understanding, comprehending plot structure, goals, and plans requires precision. Rule-based systems are appropriate for such analysis. When connecting these story elements with the details of concrete actions and specific situations, such precision isn’t needed, but broad coverage of commonsense knowledge becomes crucial. Aspire shows how rule-based systems and broad-coverage commonsense knowledge bases can work together. Being a jack-of-all-trades doesn’t mean you have to be master of none.

## Acknowledgements

This research was supported, in part, by the Air Force Office of Scientific Research, Award Number FA9550-17-1-0081.

## References

- Blass, J. A., and Forbus, K. D. 2017. Analogical Chaining with Natural Language Instruction for Commonsense Reasoning. In *AAAI*, 4357–4363.
- Chen, J., and Liu, J. 2011. Combining ConceptNet and WordNet for Word Sense Disambiguation. In *IJCNLP*, 686–694.
- Gordon, A. S. 2016. Commonsense Interpretation of Triangle Behavior. In *AAAI*, 3719–3725.
- Guha, R. V., and Lenat, D. B. 1994. Enabling agents to work together. *Commun. ACM* 37(7):126–142.
- Havasi, C.; Speer, R.; Pustejovsky, J.; and Lieberman, H. 2009. Digital Intuition: Applying Common Sense Using Dimensionality Reduction. *IEEE Intelligent systems* 24(4).
- Havasi, C.; Speer, R.; and Pustejovsky, J. 2010. Coarse Word-Sense Disambiguation Using Common Sense. In *AAAI Fall Symposium: Commonsense Knowledge*.
- Hermann, K. M.; Kocisky, T.; Grefenstette, E.; Espeholt, L.; Kay, W.; Suleyman, M.; and Blunsom, P. 2015. Teaching Machines to Read and Comprehend. In Cortes, C.; Lawrence, N. D.; Lee, D. D.; Sugiyama, M.; and Garnett, R., eds., *Advances in Neural Information Processing Systems* 28. Curran Associates, Inc. 1693–1701.
- Hobbs, J. R.; Stickel, M.; Martin, P.; and Edwards, D. 1988. Interpretation as Abduction. In *Proceedings of the 26th Annual Meeting on Association for Computational Linguistics*, 95–103. Association for Computational Linguistics.
- Holmes, D., and Winston, P. 2016. Story-enabled hypothetical reasoning. In *Advances in Cognitive Systems*.
- Katz, B. 1997. Annotating the World Wide Web using Natural Language. In *Computer-Assisted Information Searching on Internet*, 136–155. Proceedings of the 5th RIAO Conference on Computer Assisted Information Searching on the Internet.
- Kenrick, D. T.; Neuberg, S. L.; Griskevicius, V.; Becker, D. V.; and Schaller, M. 2010. Goal-driven Cognition and Functional Behavior: The Fundamental-Motives Framework. *Current Directions in Psychological Science* 19(1):63–67.
- Lai, G.; Xie, Q.; Liu, H.; Yang, Y.; and Hovy, E. 2017. RACE: Large-scale ReAding Comprehension Dataset From Examinations. *ArXiv e-prints*.
- Lascarides, A., and Asher, N. 2008. Segmented Discourse Representation Theory: Dynamic Semantics with Discourse Structure. In *Computing Meaning*. Springer. 87–124.
- Locke, E. A., and Latham, G. P. 2006. New Directions in Goal-Setting Theory. *Current Directions in Psychological Science* 15(5):265–268.
- Macbeth, J. C., and Grandic, S. 2017. Crowdsourcing a Parallel Corpus for Conceptual Analysis of Natural Language. In *Proceedings of The Fifth AAAI Conference on Human Computation and Crowdsourcing*. The Association for the Advancement of Artificial Intelligence.
- Minsky, M. L. 1986. *The Society of Mind*. Simon & Schuster Paperbacks.
- Minsky, M. L. 1991. Logical Versus Analogical or Symbolic Versus Connectionist or Neat Versus Scruffy. *AI magazine* 12(2):34.
- Mostafazadeh, N.; Roth, M.; Louis, A.; Chambers, N.; and Allen, J. F. 2017. LSDSem 2017 Shared Task: The Story Cloze Test. *LSDSem 2017* 46.
- Noss, J. M. 2016. Who Knows What? Perspective-Enabled Story Understanding. Master’s thesis, Massachusetts Institute of Technology, Cambridge, MA. <http://groups.csail.mit.edu/genesis/papers/2017%20Jessica%20Noss.pdf>.
- Rajpurkar, P.; Zhang, J.; Lopyrev, K.; and Liang, P. 2016. SQuAD: 100,000+ Questions for Machine Comprehension of Text. *arXiv preprint arXiv:1606.05250*.
- Schank, R. C., and Wilensky, R. 1977. A Goal-directed Production System for Story Understanding. *SIGART Bull.* (63):72–72.
- Schank, R. C. 1975. *Conceptual Information Processing*. New York, NY: Elsevier.
- Tandon, N.; de Melo, G.; Suchanek, F.; and Weikum, G. 2014. WebChild: Harvesting and Organizing Commonsense Knowledge from the Web. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, 523–532. ACM.
- Weston, J.; Bordes, A.; Chopra, S.; Rush, A. M.; van Merriënboer, B.; Joulin, A.; and Mikolov, T. 2015. Towards AI-Complete Question Answering: A Set of Prerequisite Toy Tasks. *arXiv preprint arXiv:1502.05698*.
- Williams, B. M. 2017. A Commonsense Approach to Story Understanding. Master’s thesis, Massachusetts Institute of Technology, Cambridge, MA. <http://groups.csail.mit.edu/genesis/papers/2017%20Bryan%20Williams.pdf>.
- Winston, P. H. 2014. The Genesis Story Understanding and Story Telling System: a 21st Century Step Toward Artificial Intelligence. Technical report, Center for Brains, Minds and Machines (CBMM).
- Young, T.; Hazarika, D.; Poria, S.; and Cambria, E. 2017. Recent Trends in Deep Learning Based Natural Language Processing. *arXiv preprint arXiv:1708.02709*.