An Exploration of Story Understanding Using Mental States in Scone

Shaan Dave

CMU-CS-22-135

August 2022

Computer Science Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Scott Fahlman, Chair
Daniel Fried

Submitted in partial fulfillment of the requirements for the degree of Master of Science.

Keywords: Scone, Artificial Intelligence, Story understanding, mental states, mental contexts, NLP

Abstract

Current AI has developed methods for working with data and producing fascinating results, but it lags behind in explicit representation and working with contradictory information, both of which are necessary for story understanding. Humans, for example, can efficiently reason about the multiple different perspectives of the people around them, even if the information held by each is not consistent. Moreover, most AIs cannot maintain contexts that inherit from one another and accurately describe a dynamic system of characters and their beliefs. Scone is a smart active memory system developed to be adept at exactly this. Scone has a virtual copy mechanism that allows it to efficiently make a clone of a world-model that differs only in a few pieces of knowledge. This allows Scone to represent different belief-states or time-states in a story. We explore what Scone is able to do through the classic fairytale of Little Red Riding Hood, illustrating its use in some examples. We explore how Scone can model characters' belief states and intentions and answer queries about them, allowing Scone to reason about and understand a story much like a human. Understanding stories, even fairytales, is an important problem for the future of AI, and Scone is a powerful tool for doing so. In addition to its contribution to AI, we hope that this work will help us gain interesting insights into human cognition through the use of Scone and its representation of mental states.

Acknowledgments

I would like to thank my advisor Scott Fahlman for taking me on as a student and introducing me to a world of concepts I would have never otherwise been exposed to. Thank you for your dedication in working on Scone and helping me throughout this process.

I also awant to thank my family, friends, and anyone, however fleeting, in my life that has ever taken the time to listen to me talk incessantly about my research.

Contents

Chapter 1

Introduction

1.0.1 Background

We show how Scone can represent different mental states and use common-sense, human-like reasoning by answering queries about stories when there are multiple characters and each has different belief states.

We show this by focusing on the Red Riding Hood story, and giving various examples of Scone interpreting the story. For example, when Red Riding Hood is on the path to her Grandmother's house and encounters the Wolf. This English portion of the story must be translated into events that can be more easily understood by Scone, which entails breaking down scenarios into much greater granularity and detail. By seeing how this can be done for the Red Riding Hood story, it will be easier to generalize my approaches for stories in general and we will see Scone's power in modeling and representing mental states and their interactions. We must remember that one of the biggest problems with AI that linguists encounter is dealing with deception and statements in which there is contradictory information (e.g. what is known should be done and what is actually planned to be done). Scone can easily handle this contradictory information using contexts and switching between them.

How are toddlers able to understand complex stories? Stories, in fact, have been around since even before civilization. They are primordial, and they have connected humans across ages, generations, and habitats. Experts agree that stories are an integral part of humanity. They can be used to communicate procedures, moral codes, traditions, and social situations. In fact, there is even merit to each of us taking some time to write stories, as it has been shown to help us deal with and figure out problems in our own lives. However, one thing about stories has always remained a mystery: how do we understand them? Even small children, barely four years of age, can understand complex ideas and reason about them through reading stories. And it just happens naturally. Many attempts have been made to understand the cognitive structures at play while a human reads and understands a story, and recent advancements in AI have allowed for novel ways for understanding the concept of a story. As we model stories through Scone and attempt to answer queries about them, demonstrating Scone's capabilities, we do not want to lose sight of the fact that this sheds light on how humans comprehend stories, how they store information, how stories affect our cognitive functioning, and, in turn, the importance of stories in the human experience.

In fact, experts note that "character states, such as desires, feelings, thoughts, and beliefs are the glue that ties the actions of the story together". And these states are noted to be a part of a causal sequence. For example in the three little pigs story, they keep running from house to house (these are their actions) but in order to understand the story the reader must realize why they are doing this, which entails understanding all of their motives and beliefs. Actually a major difference is scene in how adults describe stories versus how children 9-11 years old describe stories, in that the adults will explain the motivations of the characters while the children just summarize. Children were better able to understand multiple perspectives when making story maps, which is not too different from the kind of graphic representation we see with Scone.

To understand even the most basic of stories, a human must connect several different possibly unrelated concepts and piece together the intentions and motivations of characters through them. For example, consider this story: (1) Ada was caught vandalizing a property in New York City. (2) A cop found her and said that he has to write her up and give her a fine. (3) Since Ada was a student on financial aid, she knew that this going on her record would hurt her chances at renewing this aid. (4) However, she remembered that she was given two tickets to a Brooklyn Nets game by her school. (5) Turns out the cop went to Duke and was a big Kyrie Irving fan and wanted to impress his girlfriend, with whom he had recently gotten into a big fight. (6) The cop happily walked away and Ada didn't get a ticket that day.

What happened in this story? First let's break down all the different aspects of it that need to be understood sentence by sentence.

1. Readers must be familiar with the concept of vandalizing, and understand that it is illegal in NYC.
2. Readers must understand the implications of being written up and obtaining a fine, thus understanding the threat the cop was imposing on her.
3. Then they must make the connections that she must have a hard time paying such a fine (being on student aid), and that she would be unable to continue her education if her financial aid were not renewed.
4. Now, we switch gears a little and hold a seemingly unrelated piece of information about a basketball game in our minds, while trying to predict how this could possible connect with her getting a fine. Perhaps we even suspect a bribe coming along.
5. We learn a significant deal about the cop's motivations, and now readers understand that the cop would have a lot of reason to give into such a bribe (understanding both the implications of having an attachment to a school, a player who went to that school, and to a significant other).
6. Finally, we see the conclusion of the situation, and readers understand that the Nets tickets were given to the cop in exchange for no penalties on the offense. None of this information

is given, but is simply inferred from the cop's reaction and the consequences that befall Ada.

And this is not to mention all the extraneous information we need to have in our knowledge base already, such as what New York City is or the fact that Kyrie Irving went to Duke. Most of the aspects listed above are not explicitly mentioned in the story but rather are inferred. We even use these inferences to make further conclusions later on.

1.0.2 Motivation

While we strive to understand stories through mental contexts, and revolutionize the current approach to processing natural language, we must remember that this is only the tip of the iceburgh. In reality, we are building towards understanding more than just stories through mental contexts. See, we treat the manner in which the text is processed and converted into Scone-readable language as a black box. It is far less important how to convert a paragraph into information we can handle with Scone. Because once we have a way to funnel that information into Scone, we can generate mental contexts and reason about that. Which means that Scone generalizes much further than stories. Given a way to extract information from movies and TV shows, we could use the same exact Scone base to understand and reason about mental contexts in such media. Films would be much harder to understand than stories, since there is much greater breadth of information being received from them (for example, body language, tone of voice, etc.). Thus, we can think of our interest in understanding stories as merely an interest in a building block. We are working up to understanding more complex data such as movies. However, similarly, understanding media is simply a building block towards understanding real person to person interactions. If we want an AI that comes anywhere close to passing a Turing test, it must be able to not only extract real-time information about people's expressions and emotions and words, but also use that to create mental states and calculate with them. We are not trying to create something to read our stories for us. We are trying to create perception. We are trying to create the ability for technology to read and understand others and to reason about situations given their understanding of others.

Imagine the implications of such an AI: they could be used in a court of law, deducing information about a case and about the intentions of the accused; they could be used in negotiations, coming up with a very accurate model of what both sides desire and understanding how to come to a mutually beneficial and acceptable agreement; and they could even be used in large corporations, observing team meetings or even directly managing teams of people. If we can develop this ability to reason about mental states in computers, then as we have seen with anything we have computerized, it will far surpass our human reasoning in both ability to store information (which means in depth, accurate mental state information for a plethora of people at once) and in speed of computation. We would much rather trust these AIs and their assessments of a social situation than our own judgments, which could be inaccurate and account for far less detail than the AI.

Of course, the next step in what Scone attempts is to go beyond just listening. In fact, the AI we are talking about is nothing exceptional than the strongest active listener ever seen. It has the ability to take in information and appropriately contextualize it so that it can make logical and correct judgments about people and situations. However, usually not all information required to make an accurate judgment is present. But given accurate mental state information, this is not an issue. What do humans do when they want to learn more about an individual? They ask questions that guide them towards a deeper understanding and more accurate representation of a particular person's mental context. But such questions are not difficult to craft (at least the content of them is not) once we have mental states. It is simply a question of filling in holes. If we can develop Scone into this capacity, we will have created the most powerful socially aware AI out there, better developed than any human being. Thus we will have created and understood perception: the ability to take in information from the surrounding environment and reason about it.

Chapter 2

Mental States and Scone

2.0.1 Introduction to Mental States

Mental states are a fascinating and important concept. We hear so much nowadays about the importance of empathy. While empathy is more of a physical, biological response, such as mirror neurons firing and telling us to smile because the person in front of us is smiling, what I believe people really mean when they talk about empathy is the ability to take on the perspectives of others and reason from their point of view. This is strikingly both more valuable and more information-gathering than empathy (what good is feeling if someone is happy if you do not know why?), and it is more computationally complex than empathy (requires holding many different states at once and reasoning with them).

In fact, to address the former point that this idea is stronger than empathy, one could even argue that perspective taking, as we shall call it, even implies empathy, proving that perspective taking is at least as strong as empathy: with mental state information of another, one could reason about what they are currently thinking, and then based on that model calculate how they are feeling based on social, cultural, and biological norms that govern emotion. For example, suppose I know that my colleague is recently up for promotion at work that comes with a huge salary bonus. I also know that he recently had a baby and has been looking for a new house to move into. Because I know what he desires (getting a promotion) and his motivation (getting a new house to accommodate his newborn), once I learn that he got the promotion, I can reason that he must now be quite happy, information I may also have gained through empathy, albeit with less calculation. Perspective taking is stronger still, as I can now reason, even before he receives notice about the promotion, that my colleague would be elated if he got it and quite disappointed if he did not. It is possible to reason about future states. As we shall see, there are many other aspects of perspective taking that make it much stronger than empathy alone. However, this analysis also shows that perspective taking is more computationally complex than empathy, as while empathy requires information about the body language of an individual at a particular moment, perspective taking requires several pieces of information about one's mental state. We may even have to hold that information about a person for an extended period of time (such as weeks in the example of the colleague, for it may be a while until we learn they have received the promotion). And finally perspective taking requires significantly more steps of computation and reasoning, making it not only more space complex but also more time complex.

Let us explore the powers of perspective taking more deeply. Perspective taking has both prosocial and antisocial uses. For example, it may allow you to find the right present for your friend's birthday by helping you reason what kinds of things would make them happy. This in turn would strengthen your relationship with them, providing you with the satisfaction of their secure companionship and your vicarious happiness when they open the present. Or it can be used to save someone's life: many suicide hotlines tactfully take advantage of perspective taking in order to prevent someone from killing themselves. How do they do it? At first they are confronted with someone they have absolutely no information about, and so they actively listen to them, trying to gain as much information as possible so that they may piece together a mental state for this individual. Note once more that this is much more than feeling their pain (aka empathy), for although empathizing is certainly a major part of what supporters on such hotlines do, a far more important task is assessing why this individual is considering taking their own life. It's about their motivations, their beliefs, and their desires. In fact, many times if a supporter demonstrates to the individual in crisis that they were able to form a pretty sufficient mental model (that is, truly understand where the person is coming from), then that may be enough to calm them down. This is so alerting, as it shows just how deeply we desire for others to be sufficiently good at perspective taking so that we may be understood. Therefore, not only is the ability to perspective take useful for ourselves to have, but also it is useful to us if others have it too.

On the other hand, perspective taking may also be used to negatively influence and manipulate people. Deception, as one example, demonstrates how an understanding of another's mental state can allow us benefit at the expense of another. For example, suppose there are three individuals, Alice, Bob, and Charlie. Alice and Charlie are friends, and Bob seeks to be friends with Charlie. However, Charlie does not like Bob and complains about Bob behind his back, of which Bob recently becomes aware. So Bob tells Alice that Charlie often says negative things about her behind her back, a deception. In doing this, Bob has managed to taint the reputation of Charlie in Alice's eyes. For instance, the next time Charlie speaks negatively about Bob, Alice will imagine that Charlie also does this about her, both discounting Charlie's complaints about Bob (however valid they may be) and discounting Charlie's reliability as a friend. Alice, of course, would deem it unwise to confront Charlie about this situation, as whether or not Charlie speaks negatively of Alice behind her back, Charlie will deny the accusations. Bob has thus successfully manipulated Alice (for the time being) into being closer to him, all due to an understanding of her mental state and what she will believe.

Let us talk further about the complexity of higher order mental states and representations. That is, more than just modeling a person's beliefs and intentions. We may also model their beliefs of another person.

This is a second order mental modeling, but we do it all the time. This idea should not be confused with self referential modeling: what does my boss think of me? My friend probably thinks that I like talking a lot. My parents think that I sleep in too late. These are still just first order models that we hold about other individuals, as the "I" or "me" in each can just as easily be replaced by any other object. On the other hand, an example of a second order mental model would be "my coworker thinks that I think I am better than them." In order to make this statement, not only must we have a representation of our coworker and what they think, but also understand what the model that our coworker has of us looks like.

This can very easily be generalized to k order mental modeling, for example with A_1 thinks that A_2 thinks that A_3 thinks that $\ldots$ A_k thinks x (where x is any statement). Where we have the phrase "A_i thinks" repeated k times for i between 1 and k inclusive. As we saw in the self reference examples, the important part of this is not having k distinct subjects (in fact we need not enforce that all A_is are distinct), but having k subject-qualifier pairs.

This raises an interesting question about the distinctness of the A_is. In fact, two consecutive subject qualifier pairs cannot be indistinct. It would make no sense to say "A believes that A believes x." It is understood if A believes that A believes x, then A believes x. Proving this with contradiction is actually considerably difficult, as the proof would go as follows: suppose, for sake of contradiction, A believes that A believes x and not A believes that x. We would like to say that A believes that not x, however this is not fully realizable since it may be the case that A has no opinion about x; that is, neither A believes that x nor A believes that not x exist in the mental model of A. Even if we were to make this jump and say that A believes that not x, we could only conclude a contradiction if we first establish that there cannot be inconsistencies of this form in a person's mental model. This is a nice assumption, and we shall use it, but it is worth noting that the beliefs of the human mind are highly complex and psychologically it may be possible (for example if A views themself as a different person than themself sometimes). Perhaps even this can shed light into such psychological concepts as multiple personalities and the like.

Directly, we may say that, given consistency in mental model frameworks (and this consistency shall be useful when contradictory information is introduced and in detecting deception), if A believes that A believes that x, then when we view the mental model that A has of A, it is an exact, recursive copy of the original mental model of A (see figure below). And now by the consistency precondition, if A's model of A's beliefs has that x, then A's beliefs has that x. But further, it is simply absurd for A to have a model of A, as this will infinitely recurse, offer no valuable additional information, and, more practically, take up unreasonable (yes infinite) space, making it simply an impossibility for any human or computer to actually hold. Thus we may always consider A believes that x in place of A believes that A believes that x. Note, finally, that this is strikingly different from "A believes that A believes," as in this case we are actually seeing a first order mental model, where the second "A believes" is simply an instance of the statement x. Now we can see that in general a sequence (where + here is string concatenation") $\Sigma_{i=1}^{n} A_i$ believes $+\, x$ is equivalently

$$\Sigma_{i=1}^{k} A_i \text{ believes} + x$$

where $k <= n$ and all consecutive duplicates have been removed. This is similar to how "for all x, for all y, there exists z" is equivalent to "for all x, y there exists z" and how these would be in the same computation complexity hierarchy.

2.0.2 Background on Scone

We aim to use Scone to understand and reason about stories, but first let us understand what Scone really is.

Scone is not a complete, standalone AI. That has never been the purpose of Scone. It is a knowledge base, but in particular it is a way of managing and dealing with information, like a data structure or smart dynamic memory system. The point of Scone is not to solve the four color theorem or to find a counterexample to the Collatz conjecture—instead it focuses more on semantic and common sense reasoning, much like human reasoning. Scone can make logical deductions based on this kind of information.

One of the core features of Scone is its ability to deal with contexts. Scone can maintain several different contexts at once, with varying degrees of inheritance from one content to another. For example, we can create a context for America and create all its national laws as a part of the context. Then we can have a context for each state in America and have that inherit from the parent America context, thus obtaining all the laws associated with America. But then we may change explicit aspects of the laws of each state so that they line up with the real differences between the country and state laws. Though seemingly contradictory information is held, Scone allows for such differences. Scone can do this not only with contexts but also with any subtype, which can inherit from a parent and then efficiently add or subtract information. Finally, Scone is efficient and scalable. Scone uses marker passing algorithms, originally designed for the NETL machine (which would do significant amounts of parallel work). However, tests have shown that Scone performs well on standard computers even with millions of elements. Scone can use these marker passing algorithms to efficiently handle most common sense reasoning, just like a human. But Scone is not and was not intended to be a general theorem prover. Scone operates using Common Lisp, and can be run using Steel Bank Common Lisp, Windows-based Common Lisp systems, and CMU Common Lisp.

2.0.3 Representing Mental States

It must be a primordial question: what do we need to understand another? What kind of information must we seek to take on another person's reality? Of course, the only truly accurate way to model this would be to use the person itself as your model, as anything else would be only a mere approximation. For example, if person A were imagining the color red, and changed their thought to the color blue, any model of person A that is not person A itself would require lag time in updating its representation of what person A is currently imagining from red to blue. Thus, in the time that person A has recently switched to imagining blue and our model of person A is still under the impression that person A is imagining red, our model would be incorrect. In other words, since any model (even one that invasively measured the electrical activity of person A's and calculated at lightspeed the next thought person A has) would have some nonzero lag time in updating its representation of person A's thoughts. In this lag time, the model and person A differ in their representations of person A, demonstrating that the only truly accurate model of person A is themself. Accepting now that any model of a person and their mental states can only hope to be an approximation, the question becomes how can satisfice: can we create a model that is sufficiently quick and accurate, particularly in relation to the specific domains in which we hope to use it? For example, if we wish to use a model to identify what a person will do next, we may only need to have representations of people's beliefs, intentions, and previous actions, which is a much smaller subset of items than we would need for a complete model. This, still, is a difficult task, and is arguably one of the most important aspects of modeling another person, especially when it comes to the software of a human being (their conscious experience). In fact, if we could model what someone believes and update it in real time with the information we gain from the person and their environment, that would be a major step forward in understanding how another human functions through computation.

From our earlier reasoning about the lag time it takes for a model to update (which allows for person A and the model of person A to differ, even if for only a fraction of a second), it is clear that time is an extremely important variable in this task of understanding what another is thinking. When it comes to computation, having an understanding of continuous time is certainly not possible right now, so we must instead focus on discrete time steps. The question then arises: how do we create our discrete time steps? These would follow naturally from when we need to update a mental state (and thus would be formed by the union of all mental state changing events for all persons present in our model). Take, for example, the infamous Cheryl's Birthday logic puzzle that rose to prominence when used in a middle school examination in Singapore. It confused many people due to its simple statements and yet very interesting and subtly complex logical nature. This is the description:

Albert and Bernard just become [sic] friends with Cheryl, and they want to know when her birthday is. Cheryl gives them a list of 10 possible dates:

May 15, May 16, May 19
June 17, June 18
July 14, July 16
August 14, August 15, August 17

Cheryl then tells Albert and Bernard separately the month and the day of her birthday respectively.
Albert: I don't know when Cheryl's birthday is, but I know that Bernard doesn't know too. [sic]
Bernard: At first I don't [sic] know when Cheryl's birthday is, but I know now.
Albert: Then I also know when Cheryl's birthday is. So when is Cheryl's birthday?

Since this example involves multiple people, our time steps will be dependent on when any of Cheryl, Bernard, or Albert's thoughts and beliefs are updated. We may start with an initial state, T0, that represents all the information we know about at the start of the puzzle, such as the possible dates of Cheryl's birthday and the fact that Albert and Bernard know the month and day of her birthday, respectively. Now let us focus only on Albert. Their mental state gets updated after each line of information he tells us. Why? Because our model must gain information from Albert in order to assess if Albert's mental state has actually changed. There may be lag time, for example Albert, prior to saying that they know that Bernard doesn't know, may have already thought that and updated his mental state accordingly. The model cannot and will not know this information and can thus only update once it receives information from Albert. Let us call this new timestep T_{A1}, which is the state after Albert says his first line, and let us call the timestep after Albert's second line T_{A2}. If we were to only model Albert, these timesteps would suffice. However, we must also include a timestep for Barnard, T_{B1}, that occurs after he says his line. We then take the ordered union of these timesteps as $\{T_0, T_{A1}, T_{B1}, T_{A2}\}$ and now it is clear that these are the only points at which our model needs to update its understanding. The model is static at each time step. That is, we flow from timestep to timestep, and the only updating that needs to be done is completed at the beginning of each timestep. This structure we have now imposed upon how our model will function proves tremendously useful, as now any backtracking like in the case that information is negated or retracted (such as if a lie is told and we must remind ourselves of the model prior to this lie) can be easily done by shifting to an earlier timestate.

Chapter 3

Focus

Since there are many things I could potentially focus on, and the possibilities of what to model in Scone are nearly endless, even for a given story, and because the possibilities blow up exponentially with more events and complexity, it is important to have a clear and well-defined focus. Part of this means being clear about what my focus will NOT be, as that will help guide my research and illuminate its true purpose.

What my focus is not:

I will not get into spatial modeling. For example, I will not concern myself and the reader with details such as how many feet away do two people stand. I will not worry about where in space each person is: where they are sitting, what angle they make with the wall behind them, or what speed the raindrops are falling at. These details involve a large amount of geometrical reasoning and modeling of the environment. This is not the purpose of my project, as I wish to focus on mental states and beliefs and how they change. Thus I will make simplifications of such spatial details. What really matters is how things are in relation to each other. The fact that it is viciously pouring rain may be relevant, but the speed is not. Or the fact that two people are standing close together may be important, but the actual distance is not. In fact, this approach is very reasonable, as we do it all the time. Rarely do we pull out a measuring device and record with pinpoint accuracy how every object in a room relates to every other object. And moreover, it is ridiculous to imagine a human ever memorizing a perfect geometrical model of a room and drawing on that kind of significant amount of data in order to reason. Instead, we make simplifications and understand generalities. When we read a story, we record the major details, not dwell on particular measurements. If we do not require this level of precision in order to fully understand a story, neither does Scone, or any program for that matter. Additionally, I will not focus on complex multi-level planning. I will describe scenarios via a series of events and actions done by characters, namely the ones demonstrating high levels of change in beliefs and mental states. However, the point of this project is not the show that Scone can calculate like a chess AI. And the abilities of Scone can be equally well-demonstrated by simple events with only a few steps, and this will ease understanding of what is actually happening with the mental states. Certainly more in depth reasoning and calculation would be interesting and useful to examine, but it is not the goal of this paper. Finally, this is not an English to Scone translator. That task is in fact an entirely different project and attempting it would significantly take away from the main work being done here on mental states. Thus we will leave all code and queries in Lisp and in terms of Scone commands, with the understanding that a translation from English to Scone and vice versa can be made in the future by a sophisticated enough NLP (thus we may have some level of ambiguity between a Scone command and its English translation). I will, however, have English descriptions of all Scone commands and queries.

Chapter 4

Representing Little Red Riding Hood with Scone

I will illustrate the capabilities and prospects of Scone in working with mental contexts, specifically in the context of understanding a story. Scone, similar to human reasoning, can maintain and work with multiple different character models and reason about them, with the ability to answer queries from each perspective. The goal is to show exactly what Scone may do and what its limitations are, thus highlighting its overall strengths and weaknesses. If Scone can understand a children's story in a human-like manner, then the implications for what Scone may further understand, from more complex stories to movies to real-time human interactions, is tremendous.

4.0.1 Why Red Riding Hood

We focus particularly on the Red Riding Hood fairytale, which, though it can be traced back to pre-17th century European folk tales, is primarily attributed to the Brothers Grimm (cite). This story is perfect for initial consideration of Scone's usage. Humans under the age of five have a predilection for understanding such stories, even though they are very complex computational tasks. To date, we still struggle to create an AI capable of this feat. Then there is simply no point in tackling a more complex task—we cannot learn to walk before we can crawl. Thus we need to demonstrate an ability to understand even basic stories, such as fairy tales and fables. Moreover, the approach that Scone uses is easily generalizable: if we can accurately model and understand a story with a small number of characters, then we can certainly adapt this same model for stories with larger numbers of characters.

4.0.2 Background on Red Riding Hood

Recall Little Red Riding Hood, a fairytale often told to young children. Note that there are many versions of this story, but the one we will be using and the one we have been referring to is the Brothers Grimm version (which is considered the original).

Brief synopsis of the story:

In the fairy tale, a young girl named Little Red Riding Hood (which she is called due to the cloak that she wears), is told by her mother to visit her sick grandmother and deliver some food. Before leaving the house, her mother cautions her, mandating her to stay on the path and not veer off. Her trek to her grandmother's house involves walking through the woods. However, on her way there she encounters the Wolf, who desires to eat her. After learning that she is headed to her grandmother's house down the path, the Wolf suggests that she run deep into the woods to pick flowers for her grandmother. In this time the Wolf goes to the grandmother's house and swallows the grandmother. He then dons the grandmother's clothing and pretends to be the grandmother, as he waits for Red Riding Hood to come. When Red reaches her grandmother's house, she sees something awry with her grandmother's appearance. She comments on her grandmother's voice and eyes and, finally, her mouth. At this last remark of "what a big mouth you have," the Wolf jumps out and swallows Red too. Satisfied with his feast, the Wolf then falls asleep. Later on, a hunter comes by the area of the house and notices that something does not seem quite right. The hunter then enters the house and slashes open the Wolf with his axe. Red Riding Hood and her grandmother are both saved, and the Wolf proceeds to die.

I will demonstrate some of the queries that come up in the examples I choose and should describe some of the issues that need to be addressed and how Scone can handle these. I will create three examples in Scone based on the Red Riding Hood story (which I will describe in greater detail in the next section). I also want to highlight the overall strengths and weaknesses of Scone for representing mental contexts and changes in them to point future research about Scone and AI in the right direction. I will augment my code with diagrams to help visualize the main ideas of the transitions between contexts based on events.

Below are the two examples I will be exploring the representations of in Scone. Each example from the story will be further broken down into time steps, and we shall see how the knowledge base evolves over these time points.

4.0.3 Examples

There are two main examples we will show: Red Riding Hood first encountering the Wolf on her journey to her grandmother's house, and Red Riding Hood finding her grandmother and eventually realizing that it is the Wolf in disguise.

The first scenario we take on in this classic fairytale is when Red Riding Hood is on the path to her Grandmother's house and encounters the Wolf. This English story must be translated first into events that can be more easily understood by Scone, which entails breaking down scenarios into much greater granularity and detail. Consecutive events may include Red Riding sees the Wolf, then the Wolf hides (behind a tree), then the Wolf reveals himself. Such a simple sequence of events already requires a large amount of cognition and reasoning. From such a description of events, we can then create representations of Red Riding Hood's mental states via Scone's feature of contexts. We create a context for Red Riding Hood labeled "Red Riding Hood's belief." This represents everything that Red Riding Hood thinks is true about the world. If the Wolf reveals himself to Red Riding Hood, we create a node in the Red Riding Hood belief context that represents the Wolf. Now if we activate Red Riding Hood's belief mental context, a query asking if the Wolf is present would return true. After the Wolf hides behind the tree, we create a new active context for Red Riding Hood's belief (that inherits from her previous belief) that it no longer has such a Wolf node. Subsequent queries under this newly active mental context will result in a request for a Wolf node returning false. Thus we have successfully enabled Scone to switch between two mental contexts: Red Riding Hood's belief prior to the Wolf hiding and her belief after the Wolf hides. Scone can answer the basic question of whether or not Red Riding Hood is aware of the Wolf in either of these points. However at the same time, we as the reader of the story know the Wolf is still present after hiding, even though we can reason on behalf of Red Riding Hood and know what she thinks. We can thus add further functionality with a reality mental context, which remains unaffected by the Wolf hiding. In other words, after the event of the Wolf hiding, if we activate the reality mental context and ask if the Wolf is still present, we receive an affirmative.

We can take this further and have even more complex reasoning. Let us introduce mental states for the Wolf's beliefs. We have both what the Wolf believes and what Red Riding Hood believes that the Wolf believes. Additionally let us keep track of what the Wolf believes that Red Riding Hood believes. This latter concept of having a one agent contain beliefs about another agent can be represented by having a subcontext of the Wolf's mental state under Red Riding Hood's belief mental state. After the event that the Wolf hides, here is how we update the mental contexts that we have: Red Riding Hood believes that the Wolf is gone, and the Wolf believes that Red Riding Hood believes that the Wolf is gone. Suppose that we have a new event now, in which the Wolf's tail sticks out from behind the tree. Then we may update Red Riding Hood's belief of the Wolf's beliefs, so that now Red Riding Hood knows the Wolf is present but the Wolf still believes Red Riding Hood is oblivious. Here is a diagram to represent this.

This is a good example because there is a lot of symmetry in this diagram. Notice how after the event we see that now Red Riding Hood believes the Wolf is present and knows that the Wolf believes the Wolf is present. On the other hand, after the event the Wolf is still under the impression that RRH does not believe the Wolf is present and any queries while the Wolf's mental context is active would say that Red Riding Hood does not know the Wolf is there. I have made some simplifications in this diagram to make it more readable here, but as we can see in the red event text, this supposedly singular event may be broken up into two events: first Red Riding Hood realizes that the Wolf is present, then we have the Wolf does not notice this. However it also seems valid to assume this is one event and that the realizing action of Red Riding Hood is merely one-sided. It is interesting to note that as little information as "Red Riding Hood sees the Wolf's tail stick out" is enough for us humans to understand what occurred, whereas Scone will need more specific information. However my focus is not on how to translate the English of the story into Scone-understandable statements. Another level of complexity can be added to this: Red Riding Hood believes that the Wolf believes that Red Riding Hood believes that the Wolf is not present. This could be done by adding more levels to our hierarchy and linking a representation of Red Riding Hood's belief under the Wolf's belief that is in the context of Red Riding Hood's belief. This is already a mouthful to describe, yet this kind of reasoning is so commonplace and efficient amongst humans.

Now that we understand some of the basic components that Scone must capture and how complex even a simple scenario can become, let us focus further on elements of the story that involve trust and intentions. Specifically, why would Red leave the path when the Wolf tells her to, even though her mother forbid her from doing so? During this example, there is a lot of switching between beliefs and mental contexts.

Knowledge needed

- Where are Red and the Wolf in relation to one another
- Red comes into the interaction believing the Wolf is bad
- Red does not want to stray off the path
- The Wolf intends to trick Red and eat both her and her grandmother
- Red goes off the path to pick flowers after talking to the Wolf

Queries to answer

- Before talking to the Wolf, what does Red intend to do?
- After talking to the Wolf, what does Red intend to do?
- What is Red's perception of the Wolf at different points in time?
- What does the Wolf think that Red believes?

Finally, let us focus on the moment that Red arrives at her grandmother's house and starts to realize that her grandmother is actually the Wolf. She does this by noticing one after another that the ears, eyes, and mouth all look like that of a wolf. The main event, however, is when the Wolf jumps out of the bed at the last moment. This is the timestep at which it finally sinks in to Red that she was wrong to think that her grandmother was in the bed and she must update what she believes accordingly.

Initially, Red intends to see her grandmother, trusts the Wolf and her mother, believes that the Wolf intends to not eat her, and believes that her grandmother is in the bed.

For the Wolf, initially he intends to eat Red, knows that he himself is in the bed, and thinks that Red believes the grandmother is in the bed.

Then Wolf reveals he is in the bed, and tries to eat Red. We can imagine that the Wolf says "I want to eat you." Since this contradicts information Red has about Wolf, this cancels Red's original belief of the Wolf's intention. Similarly with the belief of grandma in bed.

Thus we see finally that Red now changes her belief of the Wolf's intention from not wanting to eat her to wanting to eat her, and changes her belief of who is in the bed as she now realizes it was not her grandmother it was the Wolf.

The Wolf, too, changes his belief of whom Red believes is in the bed, as he realizes that Red now knows the Wolf is in the bed.

Knowledge needed

- What a wolf looks like
- What a grandmother should look like
- What does a wolf talk like as opposed to a human
- A wolf can be dressed as a human

Queries to answer

- Who is in the bed when Red enters her grandmother's house?
- Who is in the bed after Red asks the Wolf each question about its appearance?
- Who does the Wolf think Red believes is in the bed at different points in time?
- When does the Wolf intend to eat Red?

What we would see in Scone

```
New-event: "Intend"
Agent := "Red"
Statement := "See grandma"
New-event: "Intend"
Agent := "Wolf"
Statement := "eat Red"
New-event: "Trust"
Agent := "Red"
Statement := "Grandma"
New-event: "Trust"
Agent := "Red"
Statement := "Wolf"
New-event: "Believe"
Agent := "Red"
Statement := "Grandma is in bed"
New-event: "Believe"
Agent := "Wolf"
Statement := "Wolf is in bed"
In-context: "Red belief of Wolf"
New-event: "Intend"
Agent := "Wolf"
Statement := "Not eat Red"
In-context: "Wolf belief of Red"
New-event: "Believe"
Agent := "Red"
Statement := "Grandma in bed"
New-event: "Reveals"
Agent := "Wolf"
```

```
Statement := "Wolf in bed"
New-event: "Not trust"
Agent := "Red"
Statement := "Wolf"
New-event: "Believe"
Agent := "Red"
Statement := "Wolf is in bed"
In-context: "Red belief of Wolf"
Event: "Intend" (stays the same)
Agent := "Wolf"
Statement := "Not eat Red"
In-context: "Wolf belief of Red"
New-event: "Believe"
Agent := "Red"
Statement := "Wolf in bed"
New-event: "Claims"
Agent := "Wolf"
Statement := "Eat Red"
New-event: "Intend"
Agent := "Red"
Statement := "See grandma"
New-event: "Intend"
Agent := "Wolf"
Statement := "eat Red"
In-context: "Red belief of Wolf"
New-event: "Intend"
Agent := "Wolf"
Statement := "Eat Red"
```

Chapter 5

Discussion

We see how Scone maintains information about each individual's mental states: what Red believes, what the Wolf believes, as well as their models of each other. Scone maintains these contexts and updates them according to the events that occur in the story, specifically events that give us as readers new information about what is going on, such as the Wolf revealing that he is actually in the bed. As soon as that event occurs, readers update their mental models of what is occuring to reflect that now Red knows the Wolf was hiding in the bed and now the Wolf knows that Red knows that. It is incredible that our human minds take such events and changes for granted. Somehow we are efficiently implementing such context switches. The way Scone does this gives us a good sense of how our own minds may be achieving it. Oftentimes, not only over the course of millennia but also more recently, our engineering methods have copied and borrowed from nature. For example, there is a Japanese transportation system that bases the way it finds efficient routes off of a type of moss that grows and spreads around the country. Why? Because due to millions of years of evolution this moss has figured out a system for determining efficient routes: it will grow outwards in several directions and then retract and take the one of least resistance. But all of our advances in technology so far have been steeped in the physical aspect of nature. There is no reason we should not look to the software of physical lifeforms to model our software upon. Scone has found a method of understanding human cognition and modeling it entirely distinct from the concept of neural networks. And it works: as a model of human cognition it is surprisingly and ostensibly robust. This has incredible implications for the future of understanding our own brains and how we perceive information as well as for understanding how to create better technology—technology that can work alongside humans, understand them, and improve their interactions with one another.