Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE 2025)

Towards a Process of Simulating Narrative Comprehension

Mica Gardone

Laboratory for Quantitative Experience Design
Kahlert School of Computing
University of Utah, Salt Lake City, UT 84106 USA
m.gardone@utah.edu

Abstract

Simulating narrative comprehension has received little attention despite potentially aiding greatly in narrative generation or aiding in designing narratives. The work must be tied to theories of mind and understanding, and be based on algorithms that are described by psychology itself. For my thesis work, I propose building a cognition-like system based on the unified narrative understanding model: Multiversionality. In order to test the algorithmic model, I present the methods by which I plan to verify the solution: 1. by testing whether updates to prior knowledge that contradict prior assumptions actually affect, slow down, comprehension, 2. by testing how knowledgeable a player is in a given genre speeds up or slows down understanding by varying the input models, and 3. by testing how different refinement strategies could encapsulate human understanding.

1 Introduction

Simulating narrative comprehension is a historical and an inciting question in artificial intelligence (AI) (McCarthy et al. 2006). Part of this discussion prompted two ways of approaching AI: a direct simulation tool and approximation through tasks. The field of AI has, broadly, accepted the stance of "*simulating* human-like abilities through *task accomplishment*" since *GPS, A Program that Simulates Human Thought* (Newell and Simon 1961). Despite the departure from "AI as a simulation tool," there has been great success and leaps between classical and statistical A.I for *machine* narrative understanding (Li et al. 2012; Cardona-Rivera et al. 2016; Lucas et al. 2024; Garoufi 2014). However, we cannot truly know if a system is approaching the fidelity of human cognition because the algorithms required are intractable (Van Rooij et al. 2024).

The Problem in Detail. Currently, should a studio or writer want to test if their experiences are being understood, it falls into one of a few categories: private test screens, focus groups, or intuition. The problem with these are the cost and/or time required to build up the requisite knowledge to make a decision. There exists little supportive software to help alleviate these costs. Specifically, no software to help pinpoint or show where experiences can go wrong *from the audience's perspective*. From an industry perspective, my work will help reduce the cost of development further and catch potential problems sooner and earlier. From an academic perspective, my work will help further human-inspired algorithms as they will be derived from psychology. In order to know if I succeeded or not, I aim to compare not just the outputs of this theoretical system, but also the *process* by which it reasons against real people.

Thus, **I propose to build a human-inspired narrative understanding agent.** This requires the agent to have precise details to simulate a *process of understanding* for narrative events. It is not enough to simply claim the solution plan is good enough. In order to claim to be human-derived, the algorithms and backing structures must be able to map onto a psychologically-valid structure (Van Rooij et al. 2024). I will be specifically focusing on narratives that require time-based dynamics. My primary effort is on developing narrative understanding, other needed components will be noted.

2 Related Work

Current methods fall into four broad categories: machine story understanding, outcome-based understanding, human-inspired machine story processing, computational cognitive modeling of story telling. Planners such as Sabre (Ware and Siler 2021) and Headspace (Sanghrajka, Young, and Thorne 2022) are not included as they are narrative generators.

Machine Story Understanding. The work of machine story understanding focuses on machines being able to derive further meaning from text, and are capable of deriving inherent events. Early work with Large Language Models typically use this as a Natural Language Processing task to demonstrate the systems' abilities over statistical modeling versions (Manning and Schütze 1999). LLMs, however, have the chance of getting story information wrong (hallucinations); their reasoning for why they are wrong is not always trustworthy (Mao et al. 2024). Further, it is incorrect to assume LLMs are capable of modeling human story understanding (Van Rooij et al. 2024).

Outcome-based Understanding. Outcome-based understanding focuses on symbolic modeling of narratives, often utilizing Q&A structures. Systems that fall into this category of work are QUEST (Graesser and Franklin 1990; Cardona-Rivera et al. 2016). Much like machine story understanding, these fail to capture human understanding as the focus is on

---

if the model can answer questions, not the process by which the story is understood.

Human-inspired Machine Story Parsing. Human-inspired machine story processing tends to focus on turning the story into parse-able chunks for a machine. Two systems that fall into this classification are Structured Event Memory (Franklin et al. 2020) and the algorithm outlined by Li et al. (2012). While closer to the goal, these processes focus on *transforming* information, not understanding.

Computational Cognitive Modeling. Computational cognition models (Niehaus and Young 2014) are computer scientists' attempts to produce *cognitively-grounded* narratives that could theoretically be understood by humans. The issue is, like the other two areas, the focus of the work is on *outputs* rather than the *process*.

In specific to Niehaus and Young (2014), the focus is on an understandable "discourse plan:" the way to *present* the plot (as opposed to plot generation, which is the most common use of planners in narrative). While it works to great effect as they describe, there are still other ways that an audience member might fail to understand the narrative not just by the discourse plan. To put it another way, their work is concerned with *if* an audience member understands the narrative, I am concerned with the *how* do they understand.

These four areas, while advancing the state of the art, do not easily (or at all, in some cases) fit the solution as discussed earlier.

3 Background

From Psychology. There are many competing models of the mind in narrative psychology, each with their own reasons for existing. These are, but not limited to, Construction-Integration (Kintsch 1988), Situation Model and Inferencing (Johnson-Laird 1983), Event Indexing (Zwaan, Langston, and Graesser 1995), and Structure Building (Gernsbacher 2013). Each of these models focuses one or a few core elements of human understanding: learning, processing, prediction, recall, et cetera. Independently, they are unable to fully capture the story understanding.

More recent work has focused on unifying these models to present different theories of mind and understanding. Of the myriad models, the model proposed by Hiskes et al. (2022) presents the most compelling model necessary to simulating narrative comprehension through *Multiversionality*; therefore, I choose to adopt this model as the basis for this work. I now summarize the theory as presented in this framework.

Constrained Expectation is how an audience member uses context and situational clues to figure out how a narrative will end. These clues can be anything *not* in the story such as the author, genre, cultural relevance, or recommendations from friends. This reduces the number of trajectories the audience member must consider during processing.

Preference Projection is an audience member's desire for an outcome, either for the narrative overall or for a specific character or group. These are entirely based on an audience member's desire that affects what steps in the narrative will be selected, and can shift as the narrative progresses.

Predictive Extrapolation is an audience member's ability to predict future events. This process mainly focuses on immediate next steps in the narrative, but can apply to longer-term narrative outcomes. According to the Multiversionality model, this process is a probabilistic process to represent the *potential* path a narrative can go.

Interpretive Extrapolation is an audience member's ability to update, modify, or replace past events in a narrative. Interpretive extrapolation is necessary to ensure predictive extrapolations and the narrative model so far are still well-founded. This process can either modify the events or orderings themselves, or modify how relevant they are in making predictive extrapolations.

To Planning. Automated planning is a well-established field rich with different models, constraints, and approaches.

Partial-Order Causal-Linked Planning is a plan-space planning technique wherein graph search is over nodes of partial plans and arcs represent refinements to those plans; the specific technique I adopt was proposed by Weld (1994). This form of refinement search is, at its core, satisfying all *open conditions* in actions (recorded as causal links) while preventing *causal threats* (where one action would clobber an established condition created by another) from potentially interrupting execution of the plan. Plan-space narrative planning is well suited to the task as a baseline; however, most plan-space work has focused on generation of narratives over understanding (Cardona-Rivera et al. 2024).

Disjunctive Temporal Planning (DTP) is a form of temporal planning that solve the Temporal Constraint Satisfaction Problem (TCSP) (Tsamardinos and Pollack 2003). DTPs differ from the classical variant as actions have *duration*, they no longer instantaneously execute. The formalism created by Schwartz and Pollack (2004), Disjunctive Temporal Partial-Order Planning (DT-POP), is a partial-order variant of the original DTP formulation. In comparison to simple temporal planning, DTPs have received less attention due to their higher complexity (Gigante et al. 2022).

Hierarchical Planning is a planning method to encode actions within other actions to better encapsulate behaviors. Most current hierarchical work is built using *hierarchical task networks* (Nau et al. 1999) (HTN), which deviates quite heavily from the POCL structure. Decompositional POCL (Young, Pollack, and Moore 1994) (DPOCL) is the POCL variant of HTNs, but has received significantly less research effort. DPOCL was originally formulated to compose speech-acts in discourse, but has also been used recently to mimic tropes in narratives (Thompson 2018) and complex camera shots (Winer and Cardona-Rivera 2018). Temporal hierarchical planning is largely planning-then-scheduling, but recent work has looked into including temporal semantics into hierarchical reasoning (Cavrel, Fiorino, and Pellier 2024; Gardone and Cardona-Rivera 2024).

4 Mapping Concepts & Defining Bounds

Next, I'll discuss the potential mappings between multiversionality and the presented automated planning systems. Some of these constructs will have clear mappings, some psychological constructs do not have clear planning analogs.

Constrained expectations and preference projection are most similar to setting up the domain and problem files of standard planning. Preference projection can utilize *preferences* (Gerevini and Long 2005) from PDDL 3. A information-rich or -poor domain and problem can also help to constrain or "blow-up" the narrative understanding space. It is my theory that modulating the domain might be like simulating narrative fluency, while modulating the problem's initial and goal states might represent how much an audience member knows or their expectations. Humans are capable of multi-level reasoning through abstraction (Crampes, Veuillez, and Ranwez 1998), such as tropes or understanding narrative beat compositions. Complex narrative information might be best served in hierarchical planning.

As I expect to be dealing with temporal narrative games, the machinery for temporal planning is obviously important. DTPs will be the basis for system development. However, current existing literature indicates something more strongly about time in narratives: when presented a choice, audiences prefer *temporally*-explained solutions over causally-refuted solutions (Kelly and Khemlani 2023).

Predictive extrapolation intuitively seems like it would be an easy mapping: use the underlying POCL refinement methods to build the narrative trajectory between the current narrative moment an the path to this point, and the goal (or preferred goal). However, this opens up a simple question: how does the audience member know what the goal is? Further, how are predictions modulated based on fluency? How far will audience members consider into the future? It may not be tenable at that moment for an audience member to fully plan how they believe a narrative goes based on current information and their own knowledge in the genre.

Interpretive extrapolation is relatively easier to discuss. There does not exist an exact mapping between what is outlined earlier and itself. The closest variant of planning is what is known as transformational planning (Müller, Kirsch, and Beetz 2007) or plan recognition (Ramırez and Geffner 2009), but that has mostly looked at through the lens of motion planning. The eventual system must be able to modify, update, or replace elements of the narrative situation model it has built and verify that all elements remain valid, or repair the model and update all preferences. Still, no system can be effectively mapped onto this aspect of cognition.

In addition to the forms of planning discussed, I also consider the underlying algorithm used. To note, empirical evidence suggests human readers want to limit the number of models (Khemlani and Johnson-Laird 2017). As such, the standard complete approach to plan-space planning is not accurate to how humans experience narratives. Beam search might be one way to simulate comprehension.

5 Evaluation

One would rightfully question how to evaluate and validate the ideas presented here. The first, and likely most obvious, question would be: what does success and failure look like? Failure is far more obvious than success: should the system not be aligning with human results during the search process, then the system clearly is not following. The degrees of success are affected by, but not limited to, how close the agent is to producing the refinement space, how close the resulting models coincide with an audience member's model, and are the inputs able to capture an audience member's knowledge of the genre, game, or work.

The next obvious question is: what tests need to be run, and what is being collected? For testing, my intuition here suggests that most tests here would have to be indirect, as stopping a testee to ask what they're thinking will invariably shift *how* they are processing information. This would mean tests would have to be conducted somewhat discreetly to not impact how the participant is processing a narrative. As such, the metrics that need to be collected to measure success are, non-exhaustively, processing speed (by introducing variably-compatible narrative events) and search comparison (the distance, or number of events, the agent is off by in comparison to humans). Processing speed can be modulated based on how model-altering the introduced event is: destructive events can cause longer processing times as models need to be remade.

6 Conclusion

In my proposed work, I seek to develop a system capable of human-inspired reasoning for narrative comprehension. This work seeks to build on established concepts in narrative psychology, using automated planning as the basis for my work. Along the way, my work will contribute to temporal planning, automated narrative sensemaking, and plan recognition. It will also help verify the algorithmic capacities described in the selected theories of mind in psychology.

Future Work. Future work can look into making the proposed model more dynamic, and therefore more human-like in a few different ways. First, as stated in Section 3, the proposed system deals simply with only a single set of preferences that remain static. In reality, longer narratives will see preferences shift overtime as the player's mental model updates. For example, a character that is once thought to be an ally to the player betrays them, leading them to seek vengeance, or hope the betraying character gets payback.

Acknowledgments

This material is based on work supported by the United States National Science Foundation (Grant #2046294).

References

Cardona-Rivera, R. E.; Jhala, A.; Porteous, J.; and Young, R. M. 2024. The story so far on narrative planning. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 34, 489–499.

Cardona-Rivera, R. E.; Price, T. W.; Winer, D. R.; and Young, R. M. 2016. Question Answering in the Context of Stories Generated by Computers. *Advances in Cognitive Systems*, 4: 227–245.

Cavrel, N.; Fiorino, H.; and Pellier, D. 2024. Extending Hierarchical Partial-Order Causal-Link Planning to Temporal Problem Solving. In *2024 IEEE 36th International Conference on Tools with Artificial Intelligence (ICTAI)*, 73–81.

Crampes, M.; Veuillez, J. P.; and Ranwez, S. 1998. Adaptive narrative abstraction. In *Proceedings of the ninth ACM conference on Hypertext and hypermedia*, 97–105.

Franklin, N. T.; Norman, K. A.; Ranganath, C.; Zacks, J. M.; and Gershman, S. J. 2020. Structured Event Memory: A Neuro-Symbolic Model of Event Cognition. *Psychological Review*.

Gardone, M.; and Cardona-Rivera, R. E. 2024. Toward Planning with Hierarchical Decompositions and Time-frames. *7th ICAPS Workshop on Hierarchical Planning (HPlan 2024)*, 7.

Garoufi, K. 2014. Planning-Based Models of Natural Language Generation: Planning-Based Models of Natural Language Generation. *Language and Linguistics Compass*, 8(1): 1–10.

Gerevini, A.; and Long, D. 2005. Plan constraints and preferences in PDDL3. Technical report, Technical Report 2005-08-07, Department of Electronics for Automation ....

Gernsbacher, M. A. 2013. *Language comprehension as structure building*. Psychology Press.

Gigante, N.; Micheli, A.; Montanari, A.; and Scala, E. 2022. Decidability and complexity of action-based temporal planning over dense time. *Artificial Intelligence*, 307: 9859–9866.

Graesser, A. C.; and Franklin, S. P. 1990. QUEST: A cognitive model of question answering. *Discourse processes*, 13(3): 279–303.

Hiskes, B.; Hicks, M.; Evola, S.; Kincaid, C.; and Breithaupt, F. 2022. Multiversionality: Considering multiple possibilities in the processing of narratives. *Review of Philosophy and Psychology*, 14(3): 1099–1124.

Johnson-Laird, P. N. 1983. *Mental models: Towards a cognitive science of language, inference, and consciousness*. 6. Harvard University Press.

Kelly, L. J.; and Khemlani, S. 2023. Temporal explanations. *Journal of Experimental Psychology: General*, 152(6): 1639.

Khemlani, S. S.; and Johnson-Laird, P. 2017. Illusions in reasoning. *Minds and Machines*, 27(1): 11–35.

Kintsch, W. 1988. The role of knowledge in discourse comprehension: a construction-integration model. *Psychological review*, 95(2): 163.

Li, B.; Lee-Urban, S.; Appling, D. S.; and Riedl, M. O. 2012. Crowdsourcing narrative intelligence. *Advances in Cognitive systems*, 2: 1–18.

Lucas, M. M.; Yang, J.; Pomeroy, J. K.; and Yang, C. C. 2024. Reasoning with large language models for medical question answering. *Journal of the American Medical Informatics Association*, 31(9): 1964–1975.

Manning, C. D.; and Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts: The MIT Press.

Mao, R.; Chen, G.; Zhang, X.; Guerin, F.; and Cambria, E. 2024. GPTEval: A Survey on Assessments of ChatGPT and GPT-4. 2308.12488.

McCarthy, J.; Minsky, M. L.; Rochester, N.; and Shannon, C. E. 2006. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI magazine*, 27(4): 12–12.

Müller, A.; Kirsch, A.; and Beetz, M. 2007. Transformational Planning for Everyday Activity. In *ICAPS*, 248–255.

Nau, D.; Cao, Y.; Lotem, A.; and Munoz-Avila, H. 1999. SHOP: simple hierarchical ordered planner. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'99, 968–973.

Newell, A.; and Simon, H. A. 1961. GPS, A Program That Simulates Human Thought. In *Computation & Intelligence: collected readings*. Rand Corporation Santa Monica, CA.

Niehaus, J.; and Young, R. M. 2014. Cognitive models of discourse comprehension for narrative generation. *Literary and Linguistic Computing*, 29(4): 561–582.

Ramırez, M.; and Geffner, H. 2009. Plan recognition as planning. In *Proceedings of the 21st international joint conference on Artifical intelligence. Morgan Kaufmann Publishers Inc*, 1778–1783.

Sanghrajka, R.; Young, R. M.; and Thorne, B. 2022. Headspace: incorporating action failure and character beliefs into narrative planning. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 18, 171–178.

Schwartz, P.; and Pollack, M. E. 2004. Planning with disjunctive temporal constraints. In *Proc. ICAPS'04 Workshop on Integrating Planning into Scheduling*, 67–74. Citeseer.

Thompson, M. 2018. *Building Abstractable Story Components with Institutions and Tropes*. Ph.D. thesis, University of Bath Bath.

Tsamardinos, I.; and Pollack, M. E. 2003. Efficient solution techniques for disjunctive temporal reasoning problems. *Artificial Intelligence*, 151(1-2): 43–89.

Van Rooij, I.; Guest, O.; Adolfi, F.; de Haan, R.; Kolokolova, A.; and Rich, P. 2024. Reclaiming AI as a Theoretical Tool for Cognitive Science. *Computational Brain & Behavior*, 1–21.

Ware, S. G.; and Siler, C. 2021. Sabre: A Narrative Planner Supporting Intention and Deep Theory of Mind. In *Proceedings of the 17th AAAI International Conference on Artificial Intelligence and Interactive Digital Entertainment*, 99–106.

Weld, D. S. 1994. An Introduction to Least Commitment Planning. *AI Magazine*, 15(4): 27–61.

Winer, D. R.; and Cardona-Rivera, R. E. 2018. A depth-balanced approach to decompositional planning for problems where hierarchical depth is requested. In *Proceedings of the 1st Hierarchical Planning Workshop at the 28th International Conference on Automated Planning and Scheduling*, 1–8.

Young, R. M.; Pollack, M. E.; and Moore, J. D. 1994. Decomposition and causality in partial-order planning. In *Proceedings of the Second International Conference on Artificial Intelligence Planning Systems, University of Chicago, Chicago, Illinois, USA, June 13-15, 1994*, 188–194.

Zwaan, R. A.; Langston, M. C.; and Graesser, A. C. 1995. The construction of situation models in narrative comprehension: An event-indexing model. *Psychological science*, 6(5): 292–297.