

UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

Curating Interest in Open Story Generation

Permalink

<https://escholarship.org/uc/item/3sm8s1zd>

Author

Behrooz, Morteza

Publication Date

2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

CURATING INTEREST IN OPEN STORY GENERATION

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTER SCIENCE

by

Morteza Behrooz

December 2019

The Dissertation of Morteza Behrooz
is approved:

Professor Arnav Jhala, Chair

Professor Katherine Isbister

Professor Sri Kurniawan

Professor Alex Pang

Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

Copyright © by

Morteza Behrooz

2019

Table of Contents

List of Figures	vi
List of Tables	vii
Abstract	ix
Dedication	xi
Acknowledgments	xii
1 Introduction	1
1.1 Contributions	8
1.2 Organization	9
I The Why	12
2 The Story of Stories	13
2.1 The Tale of All Stories	13
2.1.1 Powerful and Ubiquitous	14
2.1.2 The Root of the Power	16
2.2 The Tale of a Single Story	18
2.2.1 The Teller’s Tale	18
2.2.2 The Listener’s Tale	20
2.2.3 An Entanglement	21
2.3 Chapter Takeaways	22
3 The Joy of a Story	23
3.1 Review of Theories of Interest	24
3.1.1 Categorizations of Interest	24
3.1.2 Cognitive Interest	25

3.1.3	Interest as an Emotion	27
3.1.4	Interests Interact	27
3.2	Experiential interest	28
3.2.1	A Taxonomy of Experiential Interest	29
3.3	Heider and Simmel Experiments	30
3.3.1	A Focus on Visual Narrative Comprehension	32
3.4	Creating A Simple Animation Generator	33
3.4.1	Behavior Library	34
3.4.2	Input Values and Behavior Selection	36
3.5	Evaluating Aspects of Expanded Theory	38
3.5.1	Results and Discussions	39
3.6	Chapter Takeaways	41
4	The Making of a Storyteller	42
4.1	Story Generation	42
4.1.1	Approaches	44
4.1.2	Open Story Generation	45
4.2	Story Evaluation	47
4.2.1	The Role of Use Case on Evaluation	48
4.2.2	Evaluation Metrics and Proxy Measures	49
4.2.3	Search for Specificities	50
4.3	Chapter Takeaways	53
II	The How	54
5	Making Them Joyfully	55
5.1	A Framework	56
5.1.1	High-level Tour	56
5.1.2	Processing Input	57
5.1.3	Detecting Interesting Event Sequences	60
5.1.4	Story Generation	63
5.2	Evaluation	65
5.2.1	Raw Logs	66
5.2.2	A User Study	73
5.2.3	Results	73
5.2.4	Discussion	74
5.3	Chapter Takeaways	75
6	Telling Them Joyfully	77
6.1	Cognitive Interest as a Proxy Measure	78
6.1.1	Quantitative Estimation of Predictive Inference	78
6.1.2	Word Embeddings	79

6.1.3	Method	80
6.2	Evaluation	84
6.2.1	Study Phase I	85
6.2.2	Study Phase II	86
6.2.3	Discussions	87
6.3	Chapter Takeaways	89
 III Stories in the Wild		 91
7	Musical Stories	92
7.1	Motivation	93
7.2	Related Work	94
7.2.1	Listener Information Needs and Music Search	94
7.2.2	Voice Assistants and Music Consumption	95
7.2.3	Using Story Generation	97
7.2.4	Voice and Interactive Narrative	99
7.3	Approach and Prototype	100
7.3.1	Generating a Sequence of Segues	103
7.4	Evaluation	106
7.5	Findings	109
7.5.1	Addressing Listener Needs and Contexts	109
7.5.2	Selection of Appropriate Content for Augmentations	111
7.5.3	Appropriate Presentation of Augmentations	113
7.6	Discussions	115
7.7	Chapter Takeaways	116
8	Hedonic Shopping Stories	118
8.1	Motivation and Related Work	118
8.1.1	Hedonic Shopping Motivations	119
8.1.2	Voice Interfaces and Shopping	121
8.2	Prototype System	122
8.3	Evaluation	125
8.3.1	Setup and Procedure	125
8.3.2	Results and Discussion	126
8.4	Chapter Takeaways	130
9	Conclusions	131
9.1	Discussion	134
9.2	Recommendation for Future Work	135
 Bibliography		 137

List of Figures

3.1	A still image of the animation used in Heider and Simmel study.	31
3.2	A still image of the animation used in our generative system.	34
3.3	An overview of the generative system.	34
3.4	Still images showing snapshots of the sequence of movements in the Chase, Corner and Break-wall scenarios (in order, from top to bottom).	36
5.1	The high level architecture of the framework.	56
5.2	The interface of the system that generated the logs used in the study. Users had played a social game of rummy with a virtual agent.	65
6.1	Moving cosine similarity chart of the sample story seen in Table 6.1.	82
7.1	The system architecture of the prototype.	100
8.1	Prototype system's architecture.	123

List of Tables

3.1	Associable shape behaviors in our study.	35
3.2	Input value and compatible behavior combinations. (i, j) denotes the pair of numbers representing cognitive and experiential interests respectively.	37
3.3	The experiments in our study, conditions and input values.	39
3.4	Participant preferences (approx. percentages) and p-values from a one-tailed binomial test, for all experiments.	40
5.1	The sub-types for the two built-in generic event types.	58
5.2	The mapping between AILE parameters and sentence tokens.	64
5.3	Generic and context-specific AILEs in rummy logs. “FE” stands for Facial Expression. Sub-types, where applicable, are shown in parentheses.	66
5.4	A sample of rummy’s raw logs.	68
5.5	A sample of a rummy AIL, translation of the events shown in Table 5.4.	69
5.6	A sample full story generated for a rummy interaction.	72
5.7	Two interesting spans generated by our framework from the full story in Table 5.6.	73

5.8	Questionnaire results for every item, including a p-value from a binomial test.	74
6.1	An example story that contains a case of foreshadowing. The words in bold correspond to the dips in the chart seen in Fig. 6.1.	81
6.2	Phase I results. Ratio denotes the percentage of the stories with foreshadowing for which the presented proxy measure results in an $M > 0$	86
6.3	Phase II results. The p-value is from a one-tailed Wilcoxon Signed-Rank test.	87
7.1	Songs' metadata examples.	101
7.2	Examples for segues, their logic description, and samples for their realized text.	102
7.3	Examples for conversational augmentations.	106
7.4	Example excerpt of an augmented playlist.	107
8.1	Sample output of the system about Sneakers.	125

Abstract

Curating Interest in Open Story Generation

by

Morteza Behrooz

Stories are the most valuable currency of human communication; a fact manifested in our social lives, cultural identities and prevalent forms of entertainment. Seeking the reasons behind this level of influence takes us to a journey in human cognition. Stories are also the currency of our situated understanding of events and experiences. This deep cognitive link not only speaks to the reasons behind the influence of stories, but it also outlines specific cognitive processes involved in storytelling between humans. Awareness of these cognitive processes can enable a storyteller to tell better stories, to the point that Herman's *Storytelling and the Sciences of the Mind* [76] recommends that cognitive scientists would benefit from studying narratology and narratologists would benefit from studying cognitive science. More specifically, in [125], Keith Oatley discusses how a storyteller uses a process of theory-of-mind to tell a story tailored for the perception by a listener. Thus, when evaluating generated stories in the field of computational story generation, we also need to focus on the cognitive processes involved in the perception of stories. Crucially, the contexts in which the generated stories are used, and the approaches with which they are generated, introduce a strong influence over how this evaluation can be performed. If the semantics of the domain in which the stories are generated are known, such as it is the case in games, then a much wider set of approaches become available to both generation and evaluation of stories. However, given the advances in story generation

and potential expanded use cases of it in the future, such as interactive sociable agents, it is increasingly inefficient to assume the semantics of a domain and perform knowledge engineering before generating stories. To this end, I focus on *open story generation*, in which such a priori semantic models are not assumed. It is decidedly more challenging to generate stories through open story generation, and it is particularly more challenging to evaluate them. I believe that a focus on the perception of stories should be an integral part of this generation and evaluation, and I see open story generation to be the most compatible approach with future use cases. To this end, in this dissertation, I offer a consolidation of literature review and an evaluated expanded theory of perceived interestingness in stories. I then report on an evaluated approach to generating stories without assuming a priori semantics and using the event sequences of past interactions. Further, I will introduce an evaluation metric for the perception of stories that focuses on predictive inference and consequently cognitive interest, and show this measure to correlate with human judgment. Lastly, I will report on “stories in the wild”, the tales of two prototypes developed and evaluated in the domains of music listening and online shopping, that use story generation techniques while incorporating aspects of story interestingness theories.

For Mehri and Ali, my dear parents,
who selflessly shape who I am.

For Mohammad Hassan, my grandfather,
whose love of books, stories, and poetry shaped my interests.

And for Charles Rich, a mentor,
whose insight and guidance shaped my career.

Acknowledgments

I would like to thank my dear adviser, Arnav Jhala, for his invaluable mentorship, for his motivating me to push further at my highest, and for his keeping faith in me at my lowest. Without Arnav's willingness to pursue novel avenues of research with me, without his constant encouragement, or without his wisdom improving my work, this dissertation would not be here. More than that, without him, I would not be the researcher I am today. Thank you Arnav.

I would also like to thank my other committee members, whose valuable feedback has made this research and dissertation better. Thank you very much for your positive influence, encouragement, and support, Katherine Isbister, Sri Kurniawan, and Alex Pang.

Although he is not with us anymore, he motivated me to pursue this research and his knowledge helped make it better, when Charles Rich was part of my initial committee. Chuck brought me to graduate school and the world of research; he was a great mentor who significantly improved the direction of my life. Chuck, I miss you every day. I would like to thank Candace Sidner, my M.Sc. adviser, whose kind mentorship I am so fortunate to enjoy.

Many thanks to Michael Mateas, Jim Whitehead, Marilyn Walker, Daniel Shapiro and Richard Jullig, professors at UC Santa Cruz, who helped me with facilitating my research, their guidance, and collaboration. Without my labmates, Trevor Santarra and Michael Leece, this road would have been harder to navigate and not nearly as enjoyable. More broadly, I cannot imagine a community that fosters a better combination of intellectuality, friendship, and collaboration than the Computational Media and Computer Science departments at UC Santa Cruz; so I would like to thank peers and friends Aaron Springer, Lucas Ferreira, Adam

Summerville, James Ryan, Chandranil Chakrabortii, Paulo Gomes, Afshin Mobramaein, Ryan Compton, Jacob Garbe, Kate Compton, Stacey Mason, John Murray, Aaron Reed, Ben Samuel, Sarah Harmon, Stella Mazeika, Johnathan Pagnutti, Suzanne da Câmara, Dylan Lederle-Ensign, Batu Aytemiz, Max Kreminski, Melanie Dickinson, Cyril Focht, and Devi Acharya. Thank you, Justus Robertson, for your amazing mentorship and kind help when I needed it the most. Thank you Ian Horswill, Gillian Smith, Mark Riedl, Mike Cook, Caroline Faur, David Kent, Elisabeth Andre, and Reid Swanson for inspiring me in research and your help in my immigration process.

I would like to extend many thanks to those who taught me how to apply my knowledge in the real world; to Sarah Mennicken, a dear friend and mentor whose guidance I am so lucky to always count on; to James Skorupski, Matt MacLaurin and Anthony Topper for paving many ways for me; and to Julia Haines, Elizabeth Churchill, Henriette Cramer, Marco Zamarrato, Julie Norvaisas, Anton Zadorozhnyy, and Pree Kolari for their invaluable mentorship.

Without the support of my housemates, Kyle Liebenberg and Giordon Stark, and their willingness to chat with or play boardgames with me when I needed a break, these years would have been much harder. Many thanks to the rest of my Santa Cruz and Bay Area emotional support team; thank you, dear Beatriz, Eda, and the Laguna house; Matt, Kate, Carmen, and the Storey house; Holakou and the Liberty house; Ebrahim, Zeinab, Jordan, Ghazal, Jeremy, Chris, David, Mila, Bahareh, Mohammad, Ahmad, Mehri, Ariana, Aveena, and others.

I would like to thank my dear parents, who endured not seeing me through immigration and travel ban limitations, and who, despite being thousands of miles away, motivated me every single day. Lastly, a warm thank you to my sister Leila and my brother Mehdi who kept me sane with their kind words while I could not travel home for more than 6 years.

Chapter 1

Introduction

Story generation, as a field of research, and as a general desire to “make computers tell stories”, is older than many of the techniques used to accomplish it. A diverse set of approaches, such as planning [183], case-based reasoning [49], and machine learning [113], have been employed to accomplish one central task: to generate artifacts that encapsulate and communicate stories.

Stories have the power to entertain. The success of entire industries, such as theater, cinema, and publishing has been largely based on the appeal of storytelling. Accordingly, then, computational story generation focused mainly on generating entertaining stories. Fittingly, computer games introduced a useful crossroad where they can, in various ways, employ stories to the benefit of the player experience. This crossroad can be manifested in the overall event sequence of a game, the conversations among the game characters or between them and the player [116], the emergent stories created in a simulation-based game [148], or the interactive

narrative that can present itself in the form of a game, in which player's actions or utterances change the progression of the story [114, 168].

The story and game crossroad has been hugely beneficial to the story generation research, causing interest in the field, providing well-defined development and testing contexts, and adding to the entertainment element by enriching and enabling new forms of games. However, the appeal of storytelling in human experience is beyond such explicit forms of entertainment such as games. Storytelling is a fundamental part of our social life, and as technology aims to become more integrated with our daily lives, through ubiquitous computing, smart devices, natural interfaces, and intelligent agents, the need for new forms of story generation is becoming clear. Given that so much of our lives, interactions, and communications revolve around stories, the computational generation of stories can enrich this integration with technology. In a more explicit use case, it is hard to argue that an agent or robot has accomplished its task of being sociable if it ignores storytelling as a fundamental social behavior of humanity. For many types of interactive agents (e.g., social robots, game characters, or voice assistants) and in various contexts of interaction (e.g., entertainment, service, health care, or education) storytelling can aid or enhance interactions by increasing engagement [14], rapport [23], closeness [46], character believability [68, 140], and perceived sociability, among other factors.

Creating stories outside of games and in a way that is more suitable for open interactions with various types of interactive agents presents two important challenges:

The **first challenge** is an inability to assume the semantic models of the domain a priori. A game, unless intentionally made otherwise, defines the semantics of all artifacts, actions, and potential events inside it. This fact plays a crucial role in the process of telling a

story inside or about a game. We know, for instance, which move sequences in a game of chess are more worthy of telling a story about than others, because we know the semantic significance of the chess moves relative to the domain of chess (e.g., using information about the state of the game or the pieces involved). A story generator that cannot assume such semantics has to use other approaches to generate a good story.

Adding to this challenge is the fact that the most popular approaches taken to generating stories are the ones that necessarily involve assuming high levels of semantics. Many of the classic examples, such as Meehan's Tale-Spin [119], Minstrel [177] or Mexica [132] use planning and involve such assumption. Other and more recent examples of planning can be seen in [56, 99, 135, 142]. The same is true about story generation approaches relying on case-based and analogical reasoning [65, 127, 178]. Story Intention Graphs [52] are perhaps the best example in this regard, where characters, their goals, objects, and other artifacts of a domain are modeled in a graph which then is used to generate stories about that domain. Such a graph is a computational model of a narrative domain, and other computational models of narrative share the same characteristic of encoding domain semantics.

A different class of approaches to story generation, called *open story generation*, avoid such assumptions of domain semantics. Example approaches include crowd-sourcing, such as used in Plot Graphs [107] and the SayAnything system [167], and machine learning and text generation using neural networks [113]. While new advances are being made in this class of approaches in terms of their level of reliance on a priori semantics, they introduce new challenges of their own. Considerable effort has to be made to prepare text and story corpora to train models, and in most cases, the models or graphs created are limited to the domain of the

specific corpora. As one expands this data to include more domains, it takes an increased effort in preparing such corpus, and the generation often becomes less predictable too. For instance, control over what is generated using machine learning and neural network models is an ongoing challenge, leading to reduced applicability when it comes to generating stories about a particular situation or a set of past events and experiences.

My initial research question focuses on approaches in open story generation. I needed to investigate the viability of methods that can generate stories without assuming a priori domain semantics and are readily usable in many domains.

RQ 1: How do we generate viable stories without a priori domain semantics while staying applicable to many domains?

While a few different techniques and methods can be used to achieve this goal [107, 113, 167], it was worth investigating whether a focus on past events is a viable direction, especially since the future needs of story generation could be more situated and could heavily revolve around interactive agents and systems.

Additionally, since an ideal open story generation depends on abiding by the laws and norms of commonsense and many individual, social, and cultural contexts, avoiding explicit modeling of these phenomena would introduce a big challenge. While language models and deep learning approaches can extract some statistical knowledge from patterns in language as a whole, I wanted to focus on the remaining challenges by finding patterns in the records of past interactions. This led me to my second research question.

RQ 2: Can we create reasonably good stories from the records

of previous interactions?

I used the logs of a series of games in which an agent interacted with human players, and where, in between game moves, other interaction events such as laughter and comments occurred as well. Thus, I focused on the constrained world of such interactions where there were largely bounded expectations, and where I did not assume any semantics about the game or interaction. Besides, the techniques were developed to be generalizable to other domains.

If the local context and history of interactions are utilized, then the generated future interactions produce better stories. “Reasonably good”, however, is an inadequate evaluation metric. Hence, my **second challenge**, which is deeply related to the first one, was the difficulty of evaluation of the generated (or potential) stories. Knowing the domain semantics makes it easier to evaluate the goodness of a generated story, as described earlier in the case of a game of chess; thus, open story generation introduces a new challenge in this respect. What adds to the significance of evaluation is the implications that the use cases of a generated story have for the listener’s expectation of their quality. In a game, all of the stories are in the service of the gameplay; however, the role of stories significantly changes as they are used in other contexts, such as social interaction.

To this end, I needed to investigate what makes stories seem as interesting, outside of predefined boundaries of specific known domains. I needed to understand the dimensions along which a story, told by one human to another, is often evaluated. This led me to my third research question:

RQ 3: What is an interesting story?

This deceptively simple question is perhaps easy to answer for any one individual evaluating one story. However, storytelling is a behavior with deep roots in our cognition, and much like many other intuitive and natural behaviors, what is easy to do and evaluate for humans is often hard to computationally recreate and model. As an analogy, consider the simple behavior of nodding. While extremely natural and easy to perform, evaluate, and understand by humans, it has proven extremely difficult to computationally recreate; a considerable amount of research is done on recreating a simple nodding behavior that seems natural to people [2]. Hundreds of thousands of years of cultural and cognitive evolution have perfected our intuition-based way of evaluating such intuitive and common behavior, and so far in the case of stories, decades of psychology and cognitive science research has been dedicated to developing a better understanding of the human relationship with storytelling.

In my work, I drew from this literature to better understand story interestingness. I will introduce an extensive literature review that brings together research on this matter across many decades and presents them holistically. Moreover, I expanded on these theories and will provide new taxonomies for story interestingness. I will report on my experimental studies that aimed at evaluating aspects of said expanded theories and taxonomies.

While gaining an understanding of the qualities of the human's intuition-based evaluation of stories is the key to understanding what may contribute to better storytelling, an entirely different effort has to be undertaken in order to make those theories useful for a computational system such as a story generator. Cognitive psychology and cognitive science theories are about understanding a phenomenon and are rightly not yielded with a condition of computational viability. Moreover, avoiding an assumption of a priori semantics makes this task even more

difficult, as explained earlier. It would be particularly helpful to employ approaches such as qualitative or common-sense reasoning to evaluate interestingness, but that would necessarily limit the applicability of such approaches to new domains, due to their reliance on a priori knowledge.

Furthermore, recent advances in neural modeling of language, either in the form of Recurrent Neural Networks (RNNs) [113] or Language Models [138], have prompted the use of evaluation metrics borrowed from the related field of Natural Language Processing (NLP), such as BLEU [129] or PINC [41] scores. While linguistics and surface features of generated textual stories are certainly of significance, the overall *goodness* or *quality* of a story cannot be solely determined through them. A story that has a poor choice of words and contains grammatical errors can still be very interesting, and a story that has excellent language use and is highly coherent can still be very boring.

A useful view in understanding the evaluation of generated stories is yielded by a focus on the cognitive processes involved in the *perception* of stories. Crucially, stories do not happen in the vacuum or isolation of our choosing and are communicated, and then perceived, after they are created. Hence, if we aim to optimize the process of story generation, we often implicitly mean to optimize the instances in which the generated stories are delivered to the audience in some use case. Thus, we mean to optimize the instances of storytelling. This optimization should be reflected in our definitions of a good story and evaluation metrics, and in turn, these evaluation metrics cannot ignore the perception as a necessary part of a storytelling experience.

Researchers have often relied upon human-subjects to evaluate a story generator.

While this approach remains a gold standard, having access to automated evaluation techniques is beneficial for two reasons. Firstly, as human-subject studies can be costly and time-consuming, an automated evaluation can be performed much more frequently (e.g., for prototyping, or fine-tuning machine learning models). Secondly, operating in different contexts and domains may change the evaluation criteria, and an automated measure, if informed of such changes, could be capable of adjusting itself. For instance, in [159], researchers introduce an agent that changes its behavior based on how much rapport it has built with a long-term human companion.

To this end, my last research question was as follows.

RQ 4: How can we, without assuming domain semantics, create evaluation metrics for generated stories that focus on cognition and perception?

Throughout this research, it was crucial to test the approaches, theories, and metrics that I create with human participants. I believe that one could not effectively contribute to an area of research that depends on perception, cognition, and evaluation metrics, without conducting user studies. Nearly all of the articles published as part of this thesis work involve such studies.

1.1 Contributions

The primary contributions of this thesis are as follows.

- A comprehensive and interdisciplinary review of research on story generation and evalu-

ation with a focus on the perceived interestingness of stories.

- A computational approach to creating a validated metric for characterizing aspects of the perception of stories.
- An open-domain generation system that incorporates past experiences to generate interesting stories.

1.2 Organization

This dissertation is organized into three main parts.

Part I: The first part focuses on “The Why” of this thesis work. Storytelling, as mentioned in this introduction, is a common behavior and the effort to computationally recreate it can take various shapes. I attempt to highlight the need for a kind of story generation that focuses on interaction. Storytelling is also a natural and intuitive behavior of humans, hence, it is important to explain the need for an interdisciplinary focus on the perception of stories, understanding that process, and evaluating our understanding of it, before attempting to contribute to computational metrics of story evaluation.

- **Chapter 2: The Story of Stories** focuses on the role of stories in human life, from historic, individual and societal perspectives. I focus on the process of storytelling, the motivations of a teller, the hopes of a listener and how the teller and listener’s tales are connected: an entanglement deeply rooted in human cognition.
- **Chapter 3: The Joy of a Story** focuses on the perception of a story, the factors that

contribute to the perceived interestingness of it, and the cognitive mechanisms involved. I discuss the related research from Artificial Intelligence (AI), psychology, cognitive science, and neuroscience. I will then outline my expanded theory which consolidates and builds upon the existing knowledge, and will lastly share my evaluation of it in the form of a user study.

- **Chapter 4: The Making of a Storyteller** focuses on story generation as a field of research, outlining the related work, including both the persisting and long-held approaches such as planning, and newly emerging trends such as open story generation. I will explore the possible future of story generation, the need for a focus on evaluation, and the role and current state of automated evaluation metrics.

Part II: Motivated and informed by the research reported in part I of this dissertation, the second part reports on the new approaches that I have developed along with their evaluations in the form of user studies.

- **Chapter 5: Making Them Joyfully** discusses a new approach for generating situated stories with a focus on interaction as both the use case and the source. I will outline the method, explain the implementation of it, and share an evaluation using the logs of a system in which a virtual agent plays a social game of cards with humans.
- **Chapter 6: Telling Them Joyfully** discusses an approach for yielding an automated metric of story interestingness, without assuming a priori semantics, and with a focus on cognitive processes involved in the perception of stories. I will report on my evaluation

of this metric, discuss and compare other metrics, and discuss the compatibility of the new metric with future use cases of story generation.

Part III: Stories In the Wild The third and last part of this dissertation outlines two related case studies performed in industry, at companies Spotify and eBay, and in the domains of music listening and e-commerce respectively. I sought to investigate the role of interestingness in story generation systems employed in task-specific use cases.

- **Chapter 7: Musical Stories** outlines the first of these two studies, in which I created a system that generates a story-like experience using the metadata of music artifacts, and with a goal of semantically connecting the sequences of songs through snippets of text or voice. I will report on semi-structured interviews as an evaluation of this system and will focus on the role of story interestingness - and particularly cognitive interest - in the experience created by this prototype.
- **Chapter 8: Hedonic Shopping Stories** outlines the second of these two studies, in which I created a prototype system to generate a short story-like text or voice snippet, using the metadata of products and shopper's interests. I will again report on semi-structured interviews as an evaluation of this system and will focus on the role of story interestingness - and particularly experiential interests - in the experience created by this prototype.

Chapter 9: Lastly, and in the final chapter, I will provide a retrospective on the work. I will discuss the new directions of narrative and story generation research that this thesis highlights and lay out a myriad of extensions and new directions.

Part I

The Why

Chapter 2

The Story of Stories

The linear nature of our perception of time, the properties of our cognition and memory, and the primacy of language as our communicative tool have all contributed to the centrality of a single concept in human life: stories.

Before diving into computational story generation and evaluation metrics of stories, it is useful to study the role of storytelling in our lives, the underpinnings of its power, and the process by which storytelling is often exercised among humans.

2.1 The Tale of All Stories

Individually, we grow up listening to stories, and in turn, create and relentlessly share new ones of our own by living our lives. As societies, we use stories as effective vehicles of joy and entertainment, narrations of events and conveyors of lessons and education. And as civilizations across history, we have long utilized stories as lasting messengers of values,

beliefs, symbols, and group identities.

2.1.1 Powerful and Ubiquitous

Individual and personal storytelling, occurring among one's friends and family, is such a commonplace phenomenon that it may be easy to ignore the extent of its influence. Social storytelling plays a crucial role in shaping our interpersonal relationships and serves as a mechanism to control them. The act of telling a story to someone, the way it is told, the contents, details, timing, and even our choice of words are social carriers sending social messages and constantly redefining our relationships.

Robin Dunbar studied human conversations and found that more than two-thirds of the speaking time among people in public places is around social topics, such as personal stories and gossiping (i.e., stories about others) [51]. In order to gain a tangible example of the sheer frequency of such kinds of storytelling, we merely need to consider the most common sentence uttered in most cultures and languages after an initial greeting such as "hello": a variation of "*what is new?*". While the answer to this question does not need to be a personal and social story, it often is.

In a book titled *The Storytelling Animal: How Stories Make Us Human* [70], the author Jonathan Gottschall describes this process as follows:

"We ask our friend 'What's up?' or 'What's new?' and we begin to narrate our lives to one another, trading tales back and forth over cups of coffee or bottles of beer, unconsciously shaping and embellishing to make the tales hum. And every night, we reconvene with our loved

ones at the dinner table to share the small comedies and tragedies of our day.”

In order to better characterize social stories, we read in “*Telling the American story: A structural and cultural analysis of conversational storytelling*” [134], where Livia Polanyi focuses on the same concept under the title “conversational storytelling” and describes it as follows:

“Specific, affirmative, past time narratives which tell about a series of events which did take place at specific unique moments in a unique past time world.”

In other forms of storytelling, entire successful industries and branches of art, from theater and film to novels and games, are founded upon and flourish due to humanity’s gravitation towards stories. The joy of listening to a story precedes most of our childhood memories and this joy is carried to adulthood, where TV shows, movies, and novels are among our most beloved forms of entertainment. While our means of producing new visual, spoken, or written narratives and the ways in which we consume them rapidly evolve over time, it is never doubted that a good story, fictional or real, would lure many listeners. Much like the commonplace nature of social storytelling, our interest in crafted and entertaining stories is such a reliable factor that we tend to forget how prevalent its role is. This interest is reliable to a degree where news organizations use storytelling techniques to be efficient in their narration of real-world events: “*facts tell, stories sell*”, as a famous saying reminds us.

Stories have also proven to be an effective carrier of what civilizations have intended

to preserve; stories of a nation's values and morals, descriptors of their identities and who they are, and tales of their history and where their origins lie. Saadi's *Gulistan*¹, among countless other examples, is a book of Persian poetry in which short stories have transmitted moral values across generations for hundreds of years. As another example, and manifested in the worldwide appeal of the movie *Troy*², we see the power of *Iliad*'s storytelling about love, triumph, and conflict transcend time, language, and culture.

Stories are akin to a currency of our social life and interactions, our morals and values, our collective identity and culture, and indeed, of our commercial entertainment.

2.1.2 The Root of the Power

The interest in stories is unlikely to have persisted throughout history, across various (often disjoint) groups of people, and in many aspects of modern life solely due to being a cultural phenomenon that we are accustomed to. Indeed, it has not. What underpins such strong effect and appeal are deep and observable roots in human cognition.

Anthropologists and evolutionary biologists have been studying human's tendency to tell and listen to stories. Findings in anthropology note that storytelling might have persisted in human culture since it promotes social cohesion among groups and serves as a valuable method to pass on knowledge to future generations. Evolutionary studies, likewise, theorize that as our ancestors developed increasingly complex social relationships, stories helped them keep track of and communicate the happenings [80].

Cognitive psychologist Keith Oatley argues that fictions provide humans with a "sim-

¹[https://en.wikipedia.org/wiki/Gulistan_\(book\)](https://en.wikipedia.org/wiki/Gulistan_(book))

²[https://en.wikipedia.org/wiki/Troy_\(film\)](https://en.wikipedia.org/wiki/Troy_(film))

ulation” ground for social life. Stories can facilitate an understanding of human social circumstances that a listener has not experienced, and consequently, augment her capacity for empathy and social inference. Studies have found a link between the enjoyment of stories and better social abilities [112, 125].

Throughout human evolution, the abilities to tell, listen to, and enjoy stories became an integral part of our cognition as important and useful traits. Stories are now an integral part of the process through which we come to make sense of our own and others’ experiences [15]. More formally, stories are our cognitive tool for situated understanding and using this tool is central to the cognitive processes employed in a range of experiences from entertainment to active learning [64, 143].

Jerome Bruner’s work on narrative psychology focuses on human understanding of intentional behavior and suggests that we make sense of intentional actions by assimilating it into narrative structures [33, 34]. Focusing on a developmental perspective, there has also been specific research on how children build their worldview through creating stories of their own [54], a process that arguably has great effects on the development of their personalities.

To this end, stories appear to also be akin to a currency of the perception we have of our experiences in life. The author and educator Roger Bingham³ has experientially reached the same core lesson, and perhaps has reflected it perfectly in this simple quote:

“We tell stories in order to feel at home in the universe.”

³https://en.wikipedia.org/wiki/Roger_Bingham

2.2 The Tale of a Single Story

In this section, I will briefly focus on some of the steps involved in the telling and listening to a single story, from the motivations of telling it, to the process of perceiving it. The significance of understanding this process becomes more clear when one intends to computationally recreate it. Much like how the ubiquitous nature of storytelling may make its vast role in human life hard to detect, the intuitiveness of the process with which it happens may cause the complexities and particularities of it to go unnoticed. This focus is especially of significance as new computational approaches are enabled to generate better stories and as the surge in technological advances creates new use cases for the generated stories.

2.2.1 The Teller's Tale

As our social norms, culture, and cognitive and evolutionary tendencies drive us to tell stories, we selectively choose what stories of our experiences and observations to share with others, whom to share them with, and how to narrate them. This mechanism defines and adjusts our interpersonal and social relationships, allows us to affect how we are perceived by others, and helps us propagate the information we would like to share in groups.

Consistent with other seemingly simple aspects of storytelling, however, the process of finding out what events to tell a story about, picking the right level of abstraction and detail, and choosing the best manner of narration is an incredibly complex process to understand and even more complex to recreate. In order to help recognize this process as a distinct ability, consider the fact that individuals are able to improve this ability in themselves, to the extent

that in all of our societies, there exist professional storytellers in various capacities: novelists, scriptwriters, game creators, news reporters, and so on.

Psychologists have pointed to the role of *theory of mind* in the process of listening to a story. Humans create mental models from which they follow the events and actions inside a story. It is shown that the cognitive process of comprehension of stories shares areas of brain activation with the processing of understandings of other people [125], hinting at deep roots with the ability to engage in empathy and the theory of mind. Interestingly, researchers have studied the exact age at which this ability is developed in children. A study reported in [126] found out that the ability to follow the thoughts of imaginary characters is observable in children who are at least 5 years old, but not in those of younger ages.

Crucially, a teller of a story has experienced the listening side of stories as well; thus, through another process of theory of mind, the teller is able to control the image that is created in the audience's mind when engaging in storytelling. This image may be of our own when telling stories socially, of a fictional character when writing a novel, or of elements in a news story or an advertisement, among other things.

We experience and observe events in our daily lives and often narrate some of them to others by carefully picking the right event sequences that are worthy and interesting as a story. We decide on which details to include or leave out, and finally, we present a narration such that it causes an *interest* in the listener. Causing this interest is one of the main motivations of telling a story. At a casual level, we use cues from intuition, culture and personal techniques to maximize this interest. More professionally, authorship and storytelling skills take over to immerse the listener in a joyful experience. The boundaries between these cases, however, are

not clear since many individuals have perfected the art of social storytelling. Many individuals even gain social status from being a good teller of various types of stories.

2.2.2 The Listener's Tale

The theory of mind entangles the tales of the teller and listener of a story together. As listeners of stories, humans are distinctly able to attribute apparent behavior and perceived narratives to event sequences. A classic 1944 study by Fritz Heider and Mary-Ann Simmel [75] shows that individuals effortlessly perceive elaborate stories from the movements of abstract shapes, such as mundane rectangles and triangles. In chapter 3, I will report on a user study that recreates such abstract shape movements.

Stories also possess the power to captivate the audience and generate lasting emotions that can be inextricably tied to those of the story's characters. This is a concept referred to by psychologists as "narrative transport" [115]. In a study reported in [72], researchers have shown that prior knowledge and life experience affect the immersive experience of narrative transport. Chapter 3 also discusses the roles of such factors in the perception of interest in stories.

Such cognitive abilities of the listener, coupled with the teller's ability to use the theory of mind and *curate* interest in stories, make storytelling an effective, intuitive, and joyful shared experience. Notably, *the perception of the listener is not an independent process from the intentions of the teller in crafting the story; there exists an entanglement.*

2.2.3 An Entanglement

The tale of the listener and the tale of the teller are not separate tales coming together at the time when the story is transmitted; rather, they are deeply entangled through cognitive abilities that are encoded in humans. The perception of a story is a factor in the process of its creation, crafting and telling, through cognitive processes exercised in the mind of the teller.

This exercise may be done by the way of intuition, habit, and cognitive abilities and hence not explicitly deliberate. Alternatively, this exercise may be done by the way of careful deliberation, in which case an expert in storytelling has mastered the ways in which different stories are best perceived and crafts her storytelling in such ways, curating as much interestingness in them as possible.

It is worth noting that while stories are often personalized to fit a particular audience, context or goal of interaction, the dependence of telling and listening of stories on each other goes beyond such bespoke scenarios, and applies to any case of storytelling where one makes any attempt to tell a good story. Various levels of interest may be more subjective, contextual and situational, or more generic and rooted in more common properties of cognition.

The entanglement shines in Herman's *Storytelling and the Sciences of the Mind* [76] as it proposes that narratologists should study cognitive science and cognitive scientists should study narratology.

2.3 Chapter Takeaways

Stories are such consequential, integral and deeply integrated part of the human experience that it is hard to dissect and study them through only a process of thinking and reflection. Hence, multiple fields of research have made attempts to understand the role of stories in human life and the reasons behind their dominance. Evolutionary psychology and cognitive science have theorized about and identified reasoning that shows deep cognitive roots in the appeal of stories. Moreover, *storytelling* and *storylistening*, as acts between one or more tellers and one or more listeners, are shown to go beyond a transactional communicative phenomenon. Storytelling involves specific cognitive processes that connect and adjust the teller and listener's perception, expectations, and the subjective evaluation of the story. This hints at an entanglement between the listener and teller's roles, which is worthy of understanding in its own right and is also a deserving avenue of research for the future of computational story generation.

Chapter 3

The Joy of a Story¹

In this chapter, I will focus on the perception of stories, and in particular, the mechanisms and reasons that contribute to the perception of interest in stories.

Stories are a topic of interest across many different fields of science and given their multifaceted nature and vast role in our lives, stories are also studied from many different angles across these fields. As such, I will first report on a review of relevant literature about the perception of interest in stories across research in the fields of AI, cognitive science and psychology. Consolidation of this knowledge was a crucial aspect of developing the ability to not only understand but to expand them.

I will then report on our expanded theory of story interestingness, which is compatible with previous research and attempts to provide a basis for computational modeling of the perception of interest and a grounding for increasing interest in story generation.

¹Based on: **Behrooz, M.**, Mobramaein, A., Jhala, A., & Whitehead, J. (2018). Cognitive and Experiential Interestingness in Abstract Visual Narrative. In *Cognitive Science Society (CogSci)*.

Lastly, in this chapter, I will report on our evaluation of aspects of the said expanded theory, inspired by the classic experiments by Fritz Heider and Mary-Ann Simmel [75].

3.1 Review of Theories of Interest

Interestingness has been a classic area of research in psychology. Berlyne's theory of interestingness [19, 20], which was developed in experimentation with visual patterns, art, and music, focuses on perceptual situational interest. Berlyne considered interest to be a monotonic function of collative variables, such as novelty, complexity, uncertainty, and conflict. Later, Schank made one of the earliest attempts [152] in identifying the sources of story interestingness. With a goal of controlling inference sequences in a story understanding system, he counted the *unexpectedness* of story events, a measure of "personal relatedness" (events about those close to us), and a class of "absolute interests" (e.g., death and sex), to be the major causes of story interestingness. Absolute interests were also corroborated by other researchers under various name, such as "generically important topics" [61], or "human dramatic situations" [182].

3.1.1 Categorizations of Interest

In attempts to improve this theory, categorizing various types of interests was central to the research efforts that followed. A popular starting point in such categorization was the *source* of interest: is one interested in a stimulus because of an objective property of the stimulus, or because of predispositions in one's self? Based on attempts to answer this question,

researchers have introduced categorizations such as *individual* and *situational* interests [78], or *interestedness* and *interestingness* [62]. Beyond the question of source, Kintsch [93] proposed two types of story interestingness: “emotional” and “cognitive”. Emotional interest is created through the arousal function of certain events, and hence includes Schank’s absolute interests. Cognitive interest, on the other hand, is mostly caused by the relationship between the incoming information and background knowledge.

These categorizations led to a focus on cognitive interest, under the assumption that cognitive interest is a more universal measure and that it can more predictably attract readers and listeners to a story, regardless of the context. As a side effect, other possible sources of interest were rather neglected, and were often categorized broadly as “emotional” [92, 93], or “topic” [37] interests.

3.1.2 Cognitive Interest

Background knowledge, as previously mentioned, was introduced by Kintsch [93], along with the degree of generated uncertainty, and “postdictability” (how well the information can be meaningfully related to other sections of the story). This view shapes an inverted-U function of knowledge and uncertainty for cognitive interest, where fully-known or perfectly unknown domains are both unlikely to generate interest. However, other researchers disputed the existence of a direct causal link between background knowledge and cognitive interest. For instance, Frick [62] conducted experiments that showed background knowledge to have no direct effects on cognitive interest; instead, he concluded **a change in one’s beliefs** to be the cause of cognitive interest.

Defined as the disruption of active expectations, Mandler [110] believed **incongruity** to be the cause of cognitive interest. Under this theory, readers may implicitly assume particular schemata in every story [29], and hence, information that is incongruent with an assumed schema is considered to be a source of cognitive interest. It is worth noting that this view seems to be particularly close to Schank's unexpectedness, especially given his notion of story scripts [153].

Conceptually close to the idea of postdictability by Kintsch, a **successful resolution** of incongruity, through a process called *reconceptualization*, was believed by Iran-Nejad [82] to be the cause of cognitive interest. Iran-Nejad also associated cognitive interest with "extra cognitive operations", but this view was later deemed to be too broad by others [92].

Kim considered the **generation of inference**, which happens *as a result* of incongruity, to be more directly responsible for cognitive interest. Kim experimented with breaks in causal chains of stories to form "implicit" and "explicit" variants. As Kim points out, this theory is close to Kintsch's and Iran-Nejad's notions of postdictability and reconceptualization, but it does not require additional information in subsequent parts of the story.

Campion et al. later disputed Kim's theory in [37] and suggested that the causal breaks used in his experiments may have been a source of unexpectedness, which in turn could have caused cognitive interest. Inspired by Berlyne's notion of epistemic curiosity [21], Campion et al. focused on cognitive interest as a "motivation to know more". Through a series of experiments, they showed that cognitive interest is caused by **uncertainty**, which is in turn caused by the generation of **predictive inference**. Predictive inference, as opposed to inductive inference, involves a presumption about what will happen next in a story.

It is worth noting that neuroscience research has found observable series of activation in relevant areas of the brain during the comprehension of narrative and has associated them with predictive inference [84, 104].

3.1.3 Interest as an Emotion

Based on appraisal theories of emotion [98], Silvia points out that interest is an emotion [162]. He suggests that interest comes from two appraisals [160, 161]. First, is an evaluation of novelty-complexity, which can support previously proposed properties such as unexpectedness. Second, is an evaluation of comprehensibility or coping-potential; which involves one's belief in one's self to have the knowledge and resources to "deal with" an event (e.g., understand a complex or unexpected concept). It is worth noting that Campion et al. consider their theory to be compatible with the appraisal-based views of interest, and count the appraisal of coping-potential as a necessary condition for cognitive interest. They assert that uncertainty must both have a clear source, and be considered solvable by the audience, in order to cause cognitive interest [37].

3.1.4 Interests Interact

Studying interest in various stimuli is greatly affected by the context in which a given stimulus exists; much like day-to-day stories that people tell, which are often situated. When a story is about violence (e.g., "will the hero survive a bomb?") interests are generally higher, regardless of the particular properties in any entailed event or object. This is in line with Schank's idea of absolute interests discussed earlier. Moreover, and as another example, people's pre-

dispositions about particular common themes of life can generate or guide various kinds of interests as well (e.g., “will someone who is cheating win or lose a race?”). This idea is corroborated by other researchers [139]. As the use cases of generative storytelling systems or interactive agents increase, it is crucial to understand not only the role of interests but also the interactions between them in potentially complex situated settings. I believe that recognizing other types of interests than cognitive interests is a valuable step towards that goal.

3.2 Experiential interest

Given the blended nature of comprehension which I discussed above, one’s prior experiences and biases seem to have a significant role in the qualities of their perception of a narrative or a stimulus inside a narrative. Hence, recognizing these experiential interests, understanding them better, and studying them in conjunction with cognitive interest, can help improve our knowledge of the topic and facilitate a better computational generation of interest in stories.

I will attempt to define experiential interests as follows:

A type of interest one may hold in an external stimulus (e.g., a story) that can only be realized in the context of an audience’s natural properties, identity, preferences, prior experiences, and interactions.

I believe that experiential interests can play a central role in generating narratives for social and situated storytelling applications, especially in situated settings where memories and prior experiences are accessible.

I will now introduce a taxonomy for various types of experiential interests. These types outline a decreasing range of universality, such that type 1 is the most universal (least individualized) and type 5 is the least universal.

3.2.1 A Taxonomy of Experiential Interest

With a goal of consolidating the non-cognitive interests, I list the following taxonomy of the most common forms of experiential interests. It is worth noting that this taxonomy is not necessarily comprehensive, but an effort has been made to cover what is common in practice or known in related research.

Type 1. Instinctive Interests: called “absolute interests” by Schank [152], instinctive interests have roots in our nature. Examples: death, danger, power, sex, etc.

Type 2. Common Themes: these interests are common personal or interpersonal themes of life that happen to many individuals, and their existence and the usual circumstances around them are known by most. They might vary from culture to culture, and from generation to generation, but there also exists a great deal of consistency about many of them, across cultures and generations. Examples: being an underdog, growing up poor, being bullied in school, etc.

Type 3. Topic Interests: I use “topic interests” (in contrast to [37]) to specifically refer to *subjects* that constitute areas of general interest for individuals. Topic interests are a part of each individual’s slowly developing identity and personality. Examples: geography, sci-fi movies, fireworks, etc.

Type 4. Reminiscence: stories that are, intentionally or unintentionally, and directly or in-

directly, reminiscent of one's past, are often of significant subjective interest. Reminiscence may occur about memories of *shared experiences* as well. There is a growing body of research in cognitive and social psychology that has underlined the importance of such storytelling in self-development [118] and social relationships [3]. Examples: first dates, a road trip with an old friend.

Type 5. Implicit Familiarity: as the most personal kind of experiential interest, this type represents experiences such as *déjà vu*, in which a meaningful but not necessarily fully recognized connection between stimuli and personal memories is established [31]. Memories causing this type of interest in a stimulus are likely to be abstract and affected by emotional states. Example: a red rose reminding one of a personal experience.

A Note on “Personal Relatedness”. I believe that Schank's idea of personal relatedness is embedded into different types of experiential interest, given their range in being individualized. For instance, a premise in type 4 is the subjective significance of memories, and hence, Schank's example of cutting a child's toenails [152] fits well here.

3.3 Heider and Simmel Experiments

In a classic 1944 study [75] by Fritz Heider and Marianne Simmel, participants were shown an animation involving three moving geometrical figures, depicted in Fig. 3.1. The shapes moved around, while the “house” was stationary, with the exception of its “door”².

²To watch the animation shown to participants of the study in [75], please refer to <https://www.youtube.com/watch?v=n9TWwG4SFWQ>

The study consisted of three experiments, in which participants were asked to describe or answer questions about the animations. Nearly all of the participants of the three experiments interpreted the pictures in terms of actions of animated beings, chiefly people, who faced challenges, defended their loves, and helped the needy, among other things. Heider and Simmel's study highlighted the phenomenon of apparent behavior for the first time.

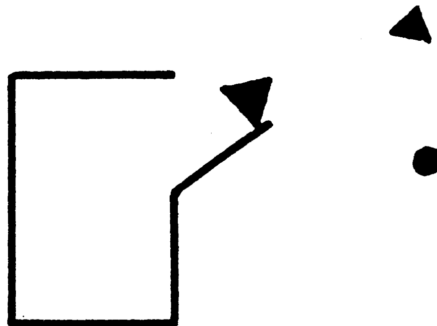


Figure 3.1: A still image of the animation used in Heider and Simmel study.

This experiment has motivated and informed researchers in many areas of study, such as social psychology [1,36,88], the psychology of art [8], and the psychology of narrative [149]. Although storytelling was not the main focus of Heider and Simmel, this study reveals the primacy of our narrative-based world view, a topic further discussed in chapter 2. Participants created stories about love, revenge, and bullying and deduced actions, goals, intentions, and personalities from simple movements of abstract shapes. The principle of attributing mental states to highly abstract stimuli has been corroborated in other studies, such as in [50].

3.3.1 A Focus on Visual Narrative Comprehension

In recent years, research in cognitive sciences has partly turned its attention towards visual narrative, such as those found in comics and films [45]. The concept of “visual attention”, which establishes a strong link with interest, is at the center of this research. Notably, *building up expectations* is assumed to be a key part of how visual narratives function in comics and movies [59]. This building up of expectations is conceptually very close to some of the theories of cognitive interest in narrative which I discussed earlier in this chapter, especially predictive inference. The parallel between these lines of research could indicate an emerging theory of visual narrative comprehension that can further and better ground a variety of lines of research on narrative, including narrative generation for entertainment and interactive agents.

Why Study Story Interestingness via Abstract Visual Cues?

Studying interest in an abstract stimulus helps our understanding of it to be free of potential nuances of various mediums and complex stimuli (e.g., games, or elaborate interactions). I believe that a theory of interest, once focused on cognitive and perceptive levels, should persist regardless of the medium in which a narrative (or indeed apparent behaviors forming a narrative) is communicated. Thus, evaluating my theory in one of the most basic forms of stimuli for comprehension, far from the biases caused by choices of particular topics, word selections, grammar use, or other language phenomena, introduces a chance to conduct a better evaluation and can confirm the depth of narrative roots in human cognition.

3.4 Creating A Simple Animation Generator

In order to study the effects of various types of interestingness in abstract visual narrative and perceived stories, and to explore the possibility of a controlled process for the generation of interest in such a setting, I designed a new system. In it, using the Unity game engine³, and based on some input values, a generator creates new animations similar to the one used in Heider and Simmel’s study [75].

The input involved two numeric values in (0, 1) range, representing a relative but quantified measure of cognitive and experiential interests. As seen in Fig. 3.2, the generated animation consisted of 3 geometrical shapes with fixed sizes and changing locations and a large fixed rectangle which included an opening, resembling a “house” and its “door”. The shapes changed their behavior, both independently and relative to each other, according to the generative system’s decision and based on the input. Other than moving, the shapes did not change in their appearance⁴.

The generator used a behavior library and a simple logic to choose various behaviors based on the pair of input values. In each instance, generating an animation involved: 1) choosing the appropriate behaviors from the library based on input values, 2) assigning behaviors to the shapes, and 3) some level of randomness in the movements of shapes (while starting and ending points were pre-programmed). An overview of this system is seen in Fig. 3.3.

³[https://en.wikipedia.org/wiki/Unity_\(game_engine\)](https://en.wikipedia.org/wiki/Unity_(game_engine))

⁴For an example of the animations generated by the system (condition 6 in our study), please refer to <https://www.youtube.com/watch?v=gB2okx77YcI>

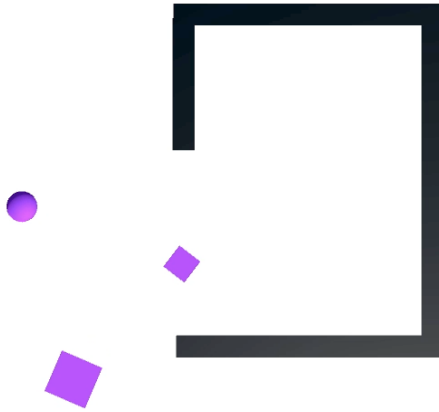


Figure 3.2: A still image of the animation used in our generative system.

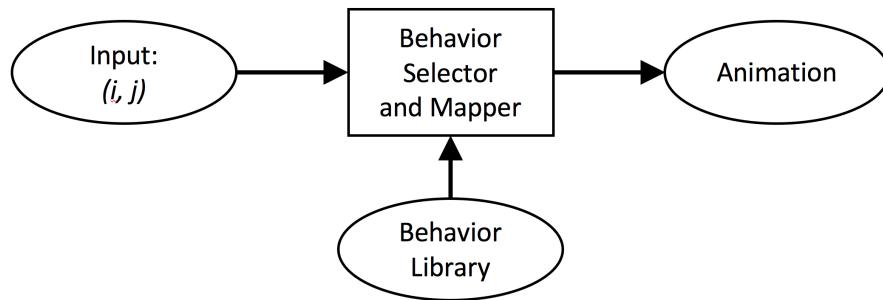


Figure 3.3: An overview of the generative system.

3.4.1 Behavior Library

In this implementation, the system supported a number of behaviors that were associated, based on the theories discussed in earlier sections, with cognitive and experiential interests. These behaviors and their descriptions are listed in Table 3.1 below. Fig. 3.4 helps visualize some of the behaviors in the library.

Table 3.1: Associable shape behaviors in our study.

Random Movements (no intended interest association): fully random movement is the most basic type of behavior of the system, during which, shapes move continuously at a constant speed and on random linear lines. They do not leave the screen, cannot cross the boundaries of the home rectangle (but can enter and exit through its door), and do not overlap with each other.

Chase (experiential): in this behavior, the larger square will randomly select one of the smaller items and chase it. The movements of the smaller item will be in random mode. The distance between the two items will change periodically to suggest tension. The chase does not end unless explicitly timed, or combined with other behaviors. The chasing behavior is a case of type 2 in experiential interests.

Corner (experiential): this behavior involves the larger square cornering one of the smaller shapes, randomly selected, in one of the three available inner corners of the home rectangle. The cornering behavior is also a case of type 2 in experiential interests.

Break-wall (cognitive): this behavior can enable a shape to ignore and cross the boundaries of the home rectangle. When followed after a period of time where walls are respected, or when observed that other shapes cannot do the same, this behavior is expected to cause cognitive interest.

Teleport (cognitive): this behavior can enable a shape to jump between any two locations of the screen, instantly, without traveling the line connecting them. When followed after a period of time without it, or when observed that other shapes cannot do the same, this behavior is expected to cause cognitive interest.

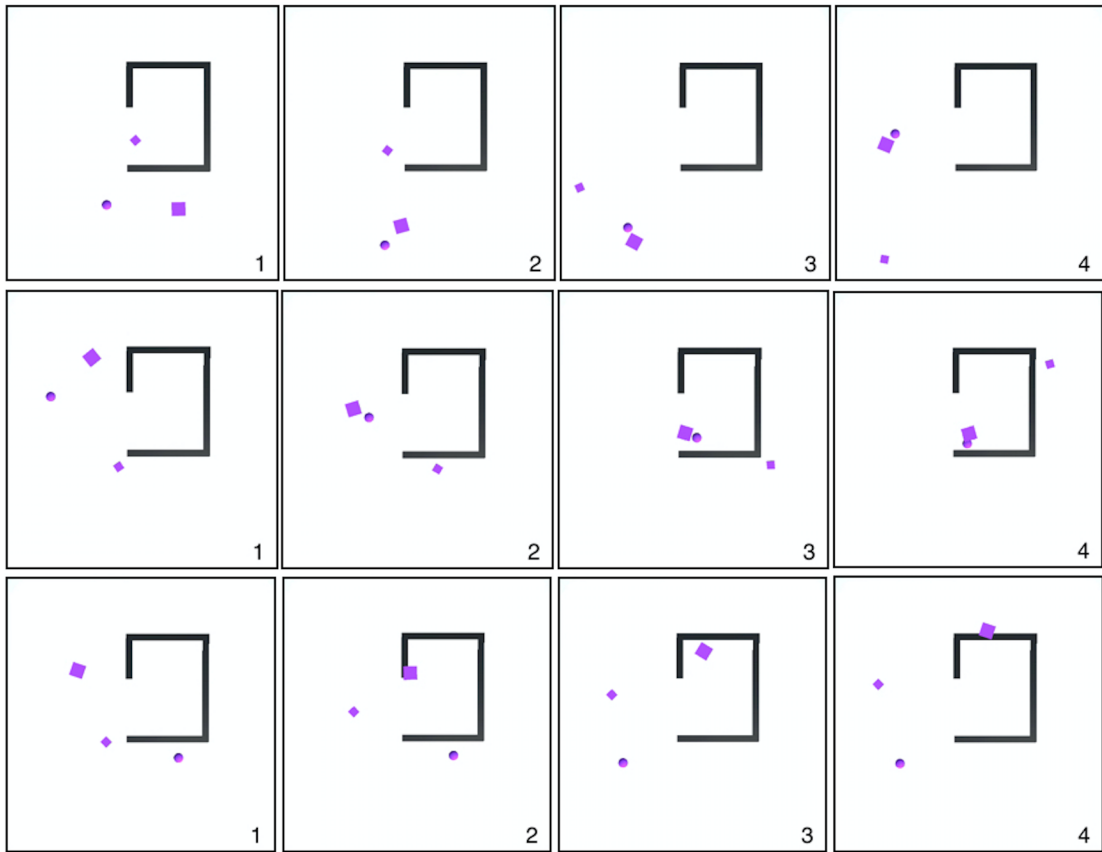


Figure 3.4: Still images showing snapshots of the sequence of movements in the Chase, Corner and Break-wall scenarios (in order, from top to bottom).

3.4.2 Input Values and Behavior Selection

In this system, the pair of input numbers could each take a value from the set $\{0, 0.5, 1\}$.

In Table 3.2, a number of possible combinations, along with compatible behaviors that may be picked by the behavior selector, are shown. In all cases, when a particular shape is not involved in any active behavior, it shows the Random Movements behavior.

Table 3.2: Input value and compatible behavior combinations. (i, j) denotes the pair of numbers representing cognitive and experiential interests respectively.

(i, j)	Behaviors
(0, 0)	With no desired level of cognitive or experiential interest, Random Movements behavior will be selected for all shapes.
(0, 0.5)	This input combination causes one of the behaviors associated with experiential interest (Chase or Corner) to be randomly selected and invoked, for an eligible shape.
(0, 1)	This input combination causes both of the behaviors associated with experiential interest to be selected, and sequentially executed (simultaneous execution would require additional shapes or altered roles). This involves a case of chasing, followed by cornering.
(0.5, 0)	This input combination causes one of the behaviors associated with cognitive interest (Break-wall or Teleport) to be randomly selected and invoked, for a random shape.
(1, 0)	This input combination causes both of the available behaviors associated with cognitive interest to be selected, and sequentially or simultaneously executed for one or multiple shapes.
(1, 1)	This maximal input combination involves Chasing, Cornering, Wall-break and Teleport. Experiential and Cognitive interest behaviors can happen simultaneously or sequentially, depending on behavior limitations; however, sequential combinations might cause less confusion.

3.5 Evaluating Aspects of Expanded Theory

Using the generation system described before, I conducted a user study to assess the perceived story interestingness of the generated animations. Such evaluation depended on the findings of Heider and Simmel [75], in that it assumes the shape movements to have a high chance of being perceived as a story.

I generated animations for 6 conditions that corresponded to using different input values, as seen in Table 3.3. Each video was between 30 to 35 seconds long; starting with 4 seconds leading time of Random Movements for all shapes and in all of the conditions. The study consisted of 8 experiments, each presenting the participants with a binary choice between two of the conditions above. The 8 experiments are seen in Table 3.3 below; each row shows the conditions ('c') compared in each experiment and the input values used for the generator. (i, j) denotes the pair of numbers representing cognitive and experiential interests respectively.

In a between-subject setting, I presented participants of each experiment with the two associated animations, in random order. I then asked the question "which animation was a more interesting story?". I recruited 60 participants for each experiment through Amazon Mechanical Turk. The participants of different experiments did not overlap.

My hypothesis was an increased perceived interestingness with greater sums of i and j . For instance, in all experiments, any intermediate condition would be perceived as more interesting than c1 (baseline), but less interesting than c6. Consequently, I further hypothesized that c3 and c5 would be perceived as more interesting than c2 and c4 respectively, and lastly, c3 as more interesting than c5.

Table 3.3: The experiments in our study, conditions and input values.

Exp.	Conditions involved	Respective input values
1	c1 vs. c3	(0,0) vs. (0,1)
2	c1 vs. c5	(0,0) vs. (1,0)
3	c1 vs. c6	(0,0) vs. (1,1)
4	c2 vs. c3	(0,0.5) vs. (0,1)
5	c4 vs. c5	(0.5,0) vs. (1,0)
6	c3 vs. c5	(0,1) vs. (1,0)
7	c3 vs. c6	(0,1) vs. (1,1)
8	c5 vs. c6	(1,0) vs. (1,1)

3.5.1 Results and Discussions

The results of the experiments are seen in Table 3.4. In all experiments, except for numbers 2 and 4, the hypotheses were confirmed with statistical significance. Experiment 6's result is borderline significant (p -value=0.046 with a one-tailed test, or 0.092 with a two-tailed), and as such, experiential interests did appear to be marginally more prominent than cognitive interests, at least in the context of this study. But crucially, the highest levels of interest were achieved when the two are combined. Overall, the results show that the generative system has been successful in creating animations (seen as visual narratives) that are perceived as interesting, along the dimensions of my proposed interest theories.

The hypothesized comparison between various quantities for cognitive interest was

Table 3.4: Participant preferences (approx. percentages) and p-values from a one-tailed binomial test, for all experiments.

Exp.	Preferences (out of 60)	p-value
1	43 (72%) for c3	< .001
2	33 (55%) for c5	.26
3	46 (77%) for c6	≪ .001
4	35 (58%) for c2	.12
5	45 (75%) for c5	≪ .001
6	37 (62%) for c3	.046
7	43 (72%) for c6	< .001
8	52 (87%) for c6	≪ .001

confirmed (experiment 5); however, this was not the case for experiential interest (experiment 4). I suspect that a longer length of Chase or Corner behaviors in condition 2 was more highlighted and perhaps dramatic for the participants than the combination of the two occurring in the same amount of time. Although, given that the null hypothesis of experiment 4 was not significantly confirmed either, it is possible that the experiential interest has such a large effect, even in small intended quantities, that differentiating between values requires more elaborated animations and behaviors.

3.6 Chapter Takeaways

Previously, in chapter 2, I discussed how the deep roots of stories in human cognition determine their dominant role in many aspects of our individual and collective lives. I also outlined how the entangled tales of the teller and listener cause attention to the perception in the process of telling. Attempting to adopt this view for computational story generation hints at a bridge that needs to be made between the processes involved in the perception of stories and the generation of them.

This chapter initially focused on the perception of stories, attempting to understand what it entails, through studying, consolidating, and expanding the relevant theories in various fields of research such as AI and cognitive science. A key factor in this process is making such theories viable for computational modeling and use by story generation systems. To this end, I introduced a consolidated and expanded theory of interest and evaluated aspects of this theory through a system generating abstract animations inspired by the Fritz Heider and Mary-Ann Simmel classic experiments on apparent behavior [75].

Chapter 4

The Making of a Storyteller

In this chapter, I will outline the approaches used in the field of story and narrative generation. I will then compare these approaches in terms of their reliance on domain semantics and the consequences of this reliance for an expansion of use cases of storytelling across domains. Besides, this reliance on semantics affects the evaluation of stories and it can make it drastically easier or harder to answer the question “is this a good story to tell?”. Thus, I will also report on an investigation of how these popular approaches lend themselves to an evaluation of stories. Lastly, I will discuss commonly used evaluation metrics from the same angle.

4.1 Story Generation

The field of narratology focuses on studying narratives, their forms, and structures [12]; for instance, separating layers in a narrative based on classes of meaning: *fabula* is the actual events in the storyworld, whereas *story* or *discourse* is the presentation of it in narration

[13, 39]. Computational narratology seeks to further formalize narrative and has been a pursuit of various fields of research, most notably, AI. Computational narratology is described as “the study of narrative from the point of view of computation and information processing” [111].

Born from this intersection of narratology and computation is story generation using AI techniques, and as the AI techniques improved over the decades, the interest in combining narratology and storytelling with AI improved as well. Several conference series and workshops are dedicated to or maintain a major focus on this area; including *Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, *Intelligent Narrative Technologies (INT)*, and *International Conference on Interactive Digital Storytelling (ICIDS)*, among others.

Most recently, advances in NLP, enabled by dramatic improvements in machine learning, have enabled the researchers in this field to shift focus from simpler “downstream” tasks such as translation and topic modeling, to higher-level and more complex goals such as story generation. Most of this effort is focused on using generative models of various types. I will discuss these approaches further in the following sections.

It is worth highlighting the branch of work on interactive narrative in this area of research. Interactive narrative is a particular form of narrative experience where the user creates or influences the progression of the unfolding drama. The main goal in this field is user immersion, where the user believes that she is an integral part of the experience [141]. Although interactive narrative has been used for education and training [147], the most common use case of it is gaming [143]; this include visual games such as *Façade* [114] and hypertext works such as *Joyce’s afternoon: a story* [86]. The recent popularity of mobile games based on interactive narrative and the availability of tools such as *Twine* [94] aiming at making the creation of in-

teractive narrative easier hints at the promise of this branch of research. More recently, Internet streaming platforms have begun to focus on interactive video¹ and audio² experiences as well.

4.1.1 Approaches

The most popular approach to generating stories has been AI planning. Classic systems such as Minstrel [177], Tale-Spin [119], Mexica [132], and many others [56, 99, 132, 135, 142] caused a flourishing of planning-based story generation systems. The crossroad of games and stories, as discussed in chapter 1, is particularly exemplified in the popularity of this approach to story generation, as planning is a useful approach in games research as well. In a more limited way, case-based and analogical reasoning has been another set of approaches used in generating stories [65, 127, 178].

What is a common factor in the approaches and examples mentioned above is that they require a priori known semantic about a domain, such as the possible set of actions for a character, in order to generate stories about that domain. Unsurprisingly, this is a good match for games, as they are usually operating in a particular domain. The world in which a game is happening and players and non-player characters act is (usually) intentionally designed in particular ways, and hence, knowing the semantics of the domain (e.g., what checkmate is in chess) is a safe assumption.

¹https://en.wikipedia.org/wiki/Black_Mirror:_Bandersnatch

²<https://www.theverge.com/2017/11/9/16602402/bbc-interactive-voice-drama-inspection-chamber-amazon-echo-alexa>

4.1.2 Open Story Generation

As I discussed in chapter 1, the advances in generative systems and the emergence of interactive agents can help the use cases of story generation to grow in such ways that, due to a need for generalizability over many unseen domains, assuming domain semantics is not viable. In chapter 5, I will provide an approach to story generation that focuses on creating stories from previous interactions without relying on semantic information about the events. I believe that such approaches are most compatible with common scenarios needed for emerging and future use cases of story generation. However, there are other existing approaches to story generation that do not rely on heavily modeling of a domain. *Open story generation* [107, 113], as this class of approaches are often called by researchers, is the problem of automatically generating a story, given some constraints such as a topic or an interaction context, without a priori semantics or manual domain knowledge engineering. Open story generation requires an intelligent system to either learn a domain model from available data about a particular domain [107, 145] or to reuse data and knowledge available from a corpus [167].

The SayAnything system [167] builds a corpus of personal stories through crowd-sourcing online blogs and uses this corpus to add statistically plausible short segments to what a human user inputs; and hence, collaboratively crafts a story with the user. Plot Graphs [107] use crowd-sourcing to find the most common sequences of events in a given domain, manifested in surface representations such as verb. The resulting graphs hold domain knowledge that can then be used to create new stories.

RNNs and Language Models have been recently used to generate stories using mod-

els trained on large corpora of stories or text [44,55,89,113,144]. Most recently, reinforcement learning has been used to control the generation of stories using neural networks by assigning a goal to this process [169]; hence, making the generation of stories more “controllable”. Although this new approach is promising, a great deal of more control is needed over the sampling of language models to make them a viable choice for generating stories for specific use cases and situations. Moreover, the interestingness in stories, as discussed in chapter 3, relies on nuanced details and it takes even more control for a sensible, applicable story to be interesting.

Story Generation from Event Sequences

As mentioned, in chapter 5, I will outline an example approach and an evaluation of it for open story generation from past event sequences. There have also been other methods developed specifically to generate narratives of past event sequences. While those methods are not necessarily cases of open story generation, their similarity to my proposed approach deserves attention.

Chess has been a particularly popular domain for exploring story generation possibilities among many researchers. The most relevant works in this area include [35] and [65]. Buchthal et al. suggest an interesting system capable of importing a chess game and applying its extracted features to one of their four story “skins”, in order to narrate them based on the dramatic elements of the chess game. This work does not produce stories directly from the game interactions, but rather uses its extracted features. Pablo Gervás takes a different approach to generating stories from chess games. This work is closer to the approach presented in chapter 5, in that it tries to find interesting event sequences in a chess game. It does so by selecting and

scoring the best “focalized” experiences within the game, which are then chosen as segments of an overall story. Gervás approach focuses on summarizing the game of chess, which differs from my work’s focus: while a good story extracted from a game does not need to represent the whole game, a summary does. Hence, Gervás approach tries to answer a question of “*how did the game go?*”, while storytelling is most related to answering a question “*did anything fun and interesting happen in the game?*”. Most importantly, both works by Buchthal et al. and Gervás described above are dependent on chess as a domain, and hence, would not be viable approaches to a generalizable story generation.

Another category of the related research has focused on deriving narrative from sport games statistics [4, 28, 97]. Such efforts are most targeted as generating “reporting style” narratives about a sports match. Such efforts are heavily conformed to the normal styles of sports news reports and hence can diverge from what a story, in general, can be. Moreover, the type of input used by such systems is often processed data that is tailored to the game events, and hence, they encode in them a great deal of semantic knowledge about the domain of the particular sports game.

4.2 Story Evaluation

All of the aforementioned approaches, especially those of open story generation (due to increased unpredictability in their generation process), would benefit from having automated evaluation metrics. In various scenarios, systems generating stories need to assess the quality of their generation. For instance, if generating stories from past events for interactive use cases,

choosing a sequence of events (and event specificities) that yields a better story needs to be an informed decision. In another example, a system that samples a Language Model to create a new fictional story for entertainment purposes also needs to have reliable evaluation factors to know which generated story is valuable to use.

4.2.1 The Role of Use Case on Evaluation

A story generator's intended use case and delivery paradigm have important implications for the constraints of the generation techniques, as well as the automated evaluation metrics it can have. As I discussed earlier, experiences such as games, which involve a mesh of different forms of entertainment, have been a great motivation for story generation. Many story generators have been developed in the context of one game, e.g. in [116], or otherwise one fixed domain, e.g. in [52]. As such, judging the quality of their generated stories can fully or partially depend on the semantics of that particular domain. This a priori knowledge can help in evaluating the quality of the generated stories, both at the fabula level (e.g., by the way of knowing the significance of the events) and at the narrative level (e.g., by a more informed choice of words or event ordering).

Future use cases of story generation, however, do not provide a guarantee in domain consistency, and in fact, the generative systems are likely to need to operate across many domains. Interactive sociable agents, for instance, often need to operate across multiple domains and contexts to increase engagement and believability. Hence, in such cases, relying on a priori domain semantics is not a viable option, neither for the generation nor for the automated evaluation processes.

4.2.2 Evaluation Metrics and Proxy Measures

Automatic assessment of the quality of the generated text is one of the main evaluation metrics in NLP and Natural Language Generation. Scores such as BLEU [129] and PINE [41] evaluate the quality of the generated text against a ground-truth source (in tasks such as translation or paraphrasing). Other scores target more generic concepts of coherence and cohesion in text [58, 71]. Perplexity is another metric for evaluating the model with which text is being generated [83], although it does not evaluate the generated text directly.

When not using the gold standard of human-subjects, story generation research - and especially open story generation - has used such language-focused metrics to evaluate the generated stories [55, 109, 113]. However, while these scores can provide some estimation of the quality of the generated text, they 1) are not suited for many newer approaches to story generation [137], and 2) do not focus on what makes stories compelling and interesting; an important consideration as discussed in previous chapters (see chapter 3).

In a relevant work, Purdy et al. introduce four quantitative story quality metrics to address these issues [137]. These measures can be used to evaluate a generated story, and are shown to correlate with the human judgment of narrative quality and enjoyment; hence, acting as “proxy measures”. These measures are:

- Correct spelling and grammar use (“**grammaticality**”),
- Linguistics-based measures of reading ease and language complexity (“**narrative productivity**”),
- Semantic similarity of adjacent sentences (“**local contextuality**”), and,

- Level of adherence to the usual ordering of events in stories; e.g., “eat” coming in a story after “order” (“**temporal ordering**”).

While the first two measures are strictly focused on the use of language, the last two focus on more semantic evaluations. Local contextuality, as defined above, uses sentence embeddings [128] to estimate semantic coherence and investigate whether sentences are relevant to each other in their progression. Temporal ordering investigates whether the verbs in a story adhere to an ordering network of precedence rules built from many stories seen before, a network similar to Plot Graphs [107].

While Purdy et al. find correlations between their proxy measures and “enjoyment” in human subjects, it is arguable that the main source of enjoyment in the perception of a story comes from finding it as *interesting*. Indeed, if a story has spelling errors, is hard to read, contains irrelevant sequences of sentences or unreasonable verbs, perceiving it would be a much less enjoyable experience than perceiving one free of those problems. However, a story that observes all such measures may still be boring and mundane. Hence, I sought to expand the notion of proxy measures and evaluation metrics to include the perceived interestingness of stories; I will discuss this new metric in chapter 6.

4.2.3 Search for Specificities

Besides forming an evaluation metric, another reason for creating proxy measures for story interestingness is the potential roles of such measures in choosing an appropriate set of specificities in a generated narrative. The two cases below exemplify such need.

Complementing Approaches Involving Generalized Concepts

In many cases, open story generation involves an explicit or implicit generalization of concepts. For instance, in [107], Plot Graphs *implicitly* generalize the specific sequences of events occurring in single stories, and form a graph that captures only the common elements across them. In another example, where generative neural networks are used to generate stories, in order to increase the chances of convergence in the model, verbs and words in the training data and story corpora are *explicitly* replaced with generalized concepts [113] using semantic word networks such as VerbNet and WordNet [120, 154]. This would result in the replacement of both of the words “car” and “automobile” with the semantic label “self-propelled vehicle.n.01”, and consequently, it becomes easier for the model to find event patterns involving either of these words. It is worth adding that in the case of Plot Graphs, in particular, generalized and most common sequences of events introduce mundane *sequences* as well as the mundane concepts.

Crucially, the narratives generated using such generalization approaches would also include generalized concepts. In the case of Plot Graphs, stories lack specificities and can be mundane without instantiation of concepts (which in turn would need domain semantics). In the case of generative neural networks, the model trained on generalized concepts is only able to generate stories involving similarly generalized concepts. It is a challenging task to replace those generalizations with specificities, as researchers report in [113].

The significance of this point lies in the interestingness and perception of stories. While a mundane story or one containing generalized concepts can be interesting, it is often the case that storytellers use the nuances in the story to make it significantly more interesting.

Having proxy measures that can help find the specificities that cause more interest may offer a solution to this problem. The next section helps provide an example scenario where such proxy measures may be helpful in a situated setting.

Picking the Right Specificity in a Situated Context

If a story generation system, for instance one used by an intelligent agent operating in the real world, attempts to build a narrative from events that have previously happened, there would be a search problem involved to choose which observations, details, or specificities (if any) should be included in the story. At a minimum, a sequence of events can be described as a mundane narrative that minimally describes the story's events. However, the inclusion of certain specificities about the elements in the story is usually what allows for authorship skills.

The *Chekhov's Gun* principle says: "every element in a story must be necessary, and irrelevant elements should be removed." On the other hand, many seemingly unnecessary parts of a telling of a story serve the particular purpose of making the narrative more interesting (e.g., through foreshadowing or red herring techniques). For instance, specifying that "the moon was shining bright" a few events before two characters (that the audience may suspect are in love) kiss for the first time, asserts a property of the moon that is (most likely) inconsequential to what happens in the story, but is nonetheless a part of what makes the telling of it more interesting.

Thus, while completely irrelevant details and specificities can violate Chekhov's Gun principle, some details and specificities, when chosen and employed in an informed and artistic way, can contribute to the interestingness of the narrative when perceived by an audience.

4.3 Chapter Takeaways

In this chapter, I briefly outlined the major different approaches taken to story generation. I focused on open story generation in which a priori domain semantics are not assumed. I further focused on the evaluation of stories, the reasons behind its growing significance, and the implications that the story generation approach carries for this evaluation. I then specifically highlighted the various evaluation metrics that can be leveraged to this end. Given the importance of perception, discussed in chapters 2 and 3, I outlined the shortcomings of the currently proposed evaluation metrics in the field. In chapter 6, I will present a new proxy measure that evaluates stories for interestingness by focusing on predictive inference and cognitive interest.

Part II

The How

Chapter 5

Making Them Joyfully¹

As I previously discussed in chapter 4, various approaches have been taken to story generation over the years, and emerging use cases require 1) flexibility over the domains they can generate stories about, and 2) support for situated systems that enable creating stories from real-world observations and experiences. To this end, this chapter explores a proposed method for creating stories from records of previous interaction, without any a priori knowledge about the semantics of the domain in which the interactions and events have occurred.

In the following sections, I will describe my proposed framework, explaining its components and their functionality. I will then report on a user study that I have conducted to evaluate this approach and will discuss its results.

¹Based on: **Behrooz, M.**, Swanson, R., & Jhala, A. (2015). Remember that time? telling interesting stories from past interactions. In *International Conference on Interactive Digital Storytelling* (pp. 93-104). Springer, Cham.



Figure 5.1: The high level architecture of the framework.

5.1 A Framework

In this section, I will introduce a system for generating interesting stories about spans of events in social interaction between a human and an intelligent agent. To this end, I designed and developed a framework with a high-level architecture depicted in Fig. 5.1. It is important to note that this framework is generic, in that it has been designed to be used for new contexts and create stories from new types of interactions with minimal changes.

5.1.1 High-level Tour

Raw interaction logs, my framework’s input, is a reasonably expected output of any external interaction systems. These logs contain information about the occurrence of various events without specifying a relation between those events and the context in which they are happening. Such external logs may contain different event types and encode them in different formats. To be able to standardize and automate the story generation process, I defined and used a notation for describing the interaction events (*AIL*, see Sec. 5.1.2). This notation allows the system to perform analyses on the events with a goal of finding sequences that are chosen to be told as a story. After finding such sequences, using a simple natural language generator (*SimpleNLG* [63]), the system can produce readable stories about the interesting spans of events in the logs. In the following sections, I will explain every component of the framework and its

functionality.

5.1.2 Processing Input

As I noted earlier, the goal of creating a new notation for describing the interaction events was to obtain a standardized format which makes it possible to analyze the event sequences statistically and to generate natural language from them. Relational logic satisfied many of these criteria while maintaining simplicity and readability. For easier reference, I will refer to this notation as *Abstract Interaction Logs* (AIL).

AIL

Inspired by Event Calculus [157], AIL consists of a series of events, each containing an *event type* and a series of parameters that provide some information about that event. AIL events are categorized within two main groups: *generic* and *context-specific*. Generic events are the ones that can happen in different contexts of interaction, such as *speech* or *facial expression*. Context-specific events are tied to a specific context in which the interaction is occurring, allowing the AIL to capture and leverage such events most efficiently. Below, is the general format of an AIL event (*AILE*) for all types:

AILE (sub, obj, type, content, context, time)

An event is context-specific if the `context` field of the AILE has a value. Moreover, `type` parameter determines the event type. Examples for event type include `speech` if the event is generic and a `meld` in the context of a rummy game. Moreover, `sub` and `obj` denote the

subject and the object for an event, which can be any of the characters or objects involved in the interaction. If the `sub` field has a value, then the event is considered to be an `action`. Furthermore, `content` allows the event to have information specific to an event, such as an utterance for a speech event, or a card number for a card game move, as much as it is extractable from the available logs. Lastly, `time` contains the time-stamp for an event. This parameter can be particularly useful if the domain involves complex temporal sequences.

Event types can have more specific and fine-grained `sub-types`. The event sub-types for the two built-in generic event types that the framework supports are shown in the Table 5.1 below.

Generic Event Type	Sub-Types
Speech	Assertives, directives, commissives, expressive, declaratives
Facial expression	Smile, laughter, gaze, nod, eyebrows-up, head-shake

Table 5.1: The sub-types for the two built-in generic event types.

In order to cover the speech sub-types most efficiently, I adopted the “illocutionary” Speech Act theory [136] and the taxonomy introduced in [156]. Below is a small description of the taxonomy:

- **Assertives:** speech acts that commit a speaker to the truth of the expressed proposition;
- **Directives:** speech acts that are to cause the hearer to take a particular action;

- **Commissives:** speech acts that commit a speaker to some future action;
- **Expressive:** speech acts that express the speaker's attitudes and emotions towards the proposition;
- **Declaratives:** speech acts that change the reality in accord with the proposition of the declaration.

The optimal coverage of the interaction events depends on the addition of sufficient context-specific events, which make it possible for the AIL to be compatible with interaction logs of different domains. Since the system does not need to know what these events mean semantically, a list of the events can be automatically created by scanning the logs as well. In a robotic task of home cleaning, for example, moving the objects is perhaps a potentially important event that can lead to interesting conversations; and in an interaction with a social virtual assistant (e.g., in a smartphone), making a phone call is likely an informative event.

Characters and Metadata

In addition to the list of AILEs, each AIL also contains a set of interaction *characters*. Each character includes at least a name and optionally a gender. The most important meta-data for an AIL is an *activity name*, which is used for creating Lead Sentences as explained in Sec. 5.1.4. Similarly to context-specific events, since the system does not rely on any a priori knowledge about the characters, the list of characters can be automatically created through scanning the logs.

Converting Raw Logs to AIL

The conversion from Raw Interaction Logs to AIL is different for various input logs and is mainly performed by text processing. After extracting the events, the framework provides easy ways to create AILE instances with various types and parameters. New AILE types can also be added to the system by inheriting from the existing types. This process needs minimal development for every new raw interaction log to ensure compatibility with log formats.

5.1.3 Detecting Interesting Event Sequences

The system's goal is to detect limited spans of a given AIL which can represent interesting event sequences in the original interaction - sequences that are considered worthy of telling a story about. The motivation behind this step is a focus on perception, as discussed in detail in chapter 2. Unexpectedness, and by proxy expectation violation, are contributing factors to cognitive interest in stories, as explained in chapter 3. Hence, without a priori semantics, this framework focuses on finding sequences that may be unexpected or could violate expectations.

Sequential Rule Mining

To this end, the system focuses on finding rare event sequences. To do so, I used a probabilistic and data mining approach. The system performs sequential rule mining by treating every AIL in a training set as one large sequence of events. Such an approach yields a series of *sequential rules*, which, as depicted below, state what series of events (*set S*) are most likely to follow a given series of events (*set P*). Sets *S* and *P* are unordered and disjoint.

$$P = (4, 5, 6) \implies S = (2, 3)$$

Since a training set is associated with a specific context, in order to find rare (or common) event sequences most efficiently, and to capture the contextual information in doing so, one needs to find sequential rules that are rare (or common) to several sequences (many AILs). Therefore, I used the *CMRules* sequential rule mining algorithm [60], which satisfies this requirement. The sequential rules yielded from this algorithm provide the system with a *support* and *confidence* number as shown below:

$$P = (4, 1, 3) \implies S = (2, 9) \text{ \#sup: .5, \#conf: .8}$$

The sequential rule above, for example, states that if events **4**, **1** and **3** are seen in any order, with a confidence of 0.8 (*conf*), they will be followed by a **2** event, and a **9** event, in some order. It also states that such a rule has a support of 0.5 (*sup*) because it appears in half of the provided event sequences.

Hence, the general sequential rule for this system follows the format below, where P and S are disjoint unordered sets of events, and *sup* and *conf* denote *support* and *confidence* as more precisely defined here.

$$(P \implies S), \text{ \#sup, \#conf}$$

- **Support:** the number of sequences that contain $P \cup S$ divided by the number of all sequences.

- **Confidence:** the number of sequences that contain $P \cup S$, divided by the number of sequences that contain P .

Using Sequential Rules to Find Rare Spans

After running the CMRules algorithm on a given AIL, the system obtains a series of sequential rules, as described earlier. The goal is to find rare spans of the interaction (event sequences) that have a higher chance of representing an interesting series of events due to unexpectedness. To do this, the system firstly considers the sequential rules with a *sup* lower than a specific threshold, which filters the original rule list significantly. The system will then sort the resulting list firstly by *conf*, and then by *sup* rates, so that the rules with lower confidence appear higher in our list. If multiple rules have equal *conf* values, then they are sorted based on their *sup* values in descending order.

Unexpectedness. In the next step, the system explores the testing set of AILs and finds matching event sequences for the list of rules previously obtained. The sorting of the rules lists is maintained in the list of extracted spans for every AIL. It is hence plausible that the extracted spans that appear higher in the list represent more interesting spans of the interaction due to being more unexpected.

Formed Expectation Violation. In an alternative approach, the system can pick the rules with the highest *conf* and *sup* (in this order), and tries to find the event sequences that violate the top rules of such list in the test set.

The system uses event `types` (and event `sub-types` where applicable) as the event types used for the sequential rule mining algorithm.

5.1.4 Story Generation

The next step in creating a story is to generate natural language based on the extracted spans of an AIL, by converting each AIL event (AILE) to a single sentence. In this process, this system uses SimpleNLG [63] which is capable of creating sentences from words and their assigned Part of Speech tags. SimpleNLG also supports changing the verb tense and the addition of adjectives and prepositional phrases.

Relating AILEs to Sentence Generation

To create a sentence from an AILE, one would first need a verb. To do this, for every AILE `sub-type`, the system keeps a verb stem that most closely describes the event or action. Afterward, it uses the story characters that are assigned to each AILE as `sub` and `obj`, and since each character contains a name string, it can obtain name strings for subject and object, where applicable. Lastly, the system will use the AILE `content` parameter as a prepositional phrase for the sentence generator.

To summarize, the conversion between an AILE's parameters and the tokens required to build a sentence is shown in Table 5.2.

AILE Parameter	Sentence Tokens
sub-type (or type)	verb
sub character's name	subject
obj character's name	object
content	propositional-phrase

Table 5.2: The mapping between AILE parameters and sentence tokens.

Creating Compound Sentences

To increase readability, one could also combine pairs of adjacent sentences with common subjects to form compound sentences (the only conjunction that is considered is “and”). This is done by a simple separately-written text processing algorithm.

Adding Lead Sentences to Story Spans

To provide a context for the story spans, if it were to be presented to the reader separately, the framework builds a single “lead sentence” to appear before a story span. This sentence plays a role similar to the role of a global orientation in the Labov-Waletzki model [96]. Lead sentences briefly inform the listener about the type of activity in the interaction. For a context-specific set of events, the system uses the activity name variable of an AIL (see Sec. 5.1.2) (e.g., “playing Rummy”), and for a generic set of events, the system uses the time of the day in which the events have been happening, if available (e.g., “In an afternoon”). In conjunction with character names, which are provided as metadata in the AIL (see Sec. 5.1.2),

lead sentences are shaped as a full sentence describing the interaction as an activity (e.g., “*During house cleaning by Jim and Karen*” or “*Once upon a time in an afternoon*”). If none of the required metadata is available, a leading phrase “*Once upon a time*” is used.

5.2 Evaluation

To evaluate this approach, I applied this framework to a specific context of social gaming. I had access to interaction logs from social gameplays between participants and a social virtual agent from my prior work (M.Sc. thesis [17]), with the related interface shown in Fig. 5.2. In the following sections, I will describe the specific interaction logs and context-specific events used for this study, explain the evaluation, and present and discuss the results.



Figure 5.2: The interface of the system that generated the logs used in the study. Users had played a social game of rummy with a virtual agent.

5.2.1 Raw Logs

The gameplay logs database contained 15 cases of interaction logs happening during a rummy game. The games were played between human players and a social virtual agent. The agent was capable of making verbal comments about the gameplay and showing facial expressions related to the events in the game [17]. The logs contained time-stamped records of the following events:

- User’s game move (Draw, Discard, Layoff, Meld)
- Agent’s game move (Draw, Discard, Layoff, Meld)
- User’s smile and laughter
- User’s comment on a move or response to an agent’s comment
- Agent’s comments on a move or response to a user’s comment

These events are translated into generic and context-specific AILEs seen in Table 5.3 below.

Generic	Speech (expressive), FE (smile), FE (laughter)
Context-Specific	Draw, Discard, Meld, Layoff

Table 5.3: Generic and context-specific AILEs in rummy logs. “FE” stands for Facial Expression. Sub-types, where applicable, are shown in parentheses.

A sample section of rummy’s raw logs, along with the respective AIL translation, can be found in Tables 5.4 and 5.5. Null values are used when a parameter is not applicable.


```
Feb 20, 2014 11:15:50 AM edu.wpi.always.srummy.SrummyClient
receivedMessage
INFO: agent laid off: 0.6
Feb 20, 2014 11:15:56 AM edu.wpi.always.srummy.SrummyClient
receivedMessage
INFO: agent discarded: 0.4
Feb 20, 2014 11:16:00 AM edu.wpi.always.srummy.StartGamingSequence
$AgentComments$1 run
INFO: Agent comment: i am finishing my cards
Feb 20, 2014 11:16:04 AM edu.wpi.always.srummy.StartGamingSequence
$HumanResponds$1 run
INFO: User selected response: not so soon
Feb 20, 2014 11:16:05 AM edu.wpi.always.srummy.SrummySchema
logTheHappinessValue
INFO: Happiness value: 1
Feb 20, 2014 11:16:09 AM edu.wpi.always.srummy.SrummySchema
logTheHappinessValue
INFO: Happiness value: 0
Feb 20, 2014 11:16:17 AM edu.wpi.always.srummy.SrummyClient
receivedMessage
INFO: User melded: 0.8
```

```

Feb 20, 2014 11:16:17 AM edu.wpi.always.srummy.StartGamingSequence
$gameOverDialogueByHuman init
INFO: Game Over, human wins.
Feb 20, 2014 11:16:18 AM edu.wpi.always.srummy.SrummySchema
logTheHappinessValue
INFO: Happiness value: 0
Feb 20, 2014 11:16:19 AM edu.wpi.always.srummy.SrummySchema
logTheHappinessValue
INFO: Happiness value: 25
Feb 20, 2014 11:16:19 AM edu.wpi.always.srummy.SrummySchema
logTheHappinessValue
INFO: Happiness value: 78
Feb 20, 2014 11:16:27 AM edu.wpi.always.srummy.StartGamingSequence
$gameOverDialogueByHuman$1 run
INFO: Game Over - comment chosen by user: I finally win. Good game.

```

Table 5.4: A sample of rummy’s raw logs.

```

LayOffEventAILI(Karen, card, null, null , rummy, 11:15:50)
DiscardEventAILI(Karen, card, null, null , rummy, 11:15:56)
SpeechAILI(Karen, User, EXPRESSIVE, yes, i am finishing my cards, null,
11:16:03)

```

ResponseAILI (User, Karen, EXPRESSIVE, not so soon, rummy, 11:16:04)
MeldEventAILI (User, card, null, null , rummy, 11:16:17)
FacialExpressionAILI (User, null, SMILE, null , null, 11:16:19)
GameOverAILI (game, null, null, null , rummy, 11:16:27)
SpeechAILI (User, Karen, EXPRESSIVE, i finally win. good game., null, 11:16:27)
FacialExpressionAILI (User, null, LAUGHTER, null , null, 11:16:28)
FacialExpressionAILI (User, null, SMILE, null , null, 11:16:29)
FacialExpressionAILI (User, null, LAUGHTER, null , null, 11:16:29)

Table 5.5: A sample of a rummy AIL, translation of the events shown in Table 5.4.

The framework created a model as described in previous sections by scanning logs of 14 instances of this social gameplay. I then used the remaining one log as a test case. This selection was done randomly. Table 5.6 shows the full rummy story of this test case, consisting of all the events in the interaction, and translated into natural language. Table 5.7 shows the interesting spans of the same log, selected by the framework. In the first span, it is seen that the agent (Karen) seems to be on a streak of strong moves, and the user is complementing Karen. In the second span, Karen brags about finishing her cards after playing a layoff move. User tells Karen “not so soon”, and plays a finishing meld move and wins the game. User then reacts to the situation by smiling and talking about it.

There once was an agent named Karen who was playing rummy with a human named user.

Karen melded card. Next, Karen discarded card.

Karen told User how's your hand over there?.

Afterwards, User responded to Karen good!.

User discarded card.

Subsequently, Karen told User i should say, you do play very well.

User responded to Karen thank you, you do too.

Karen discarded card.

User laughed. Next, User smiled. After that, User laughed.

After which, User discarded card.

Karen discarded card.

After that, User stopped laughing.

User melded card. User laughed. User smiled.

Afterwards, User discarded card.

Karen laid off card. Karen discarded card.

User told Karen nice. After which, User smiled.

After which, User discarded card.

Karen laid off card. Karen discarded card.

User told Karen nice again.

User laid off card. After that, User discarded card.

User told Karen found a lay off, yes!. Karen responded to User.

After that, Karen discarded card.

After which, Karen told User i am in this to win.

Afterwards, User responded to Karen well do your best.

User discarded card. After that, Karen discarded card.

User told Karen you do play very well.

Karen responded to User good job.

User discarded card.

Karen told User humans are somewhat intelligent but robots are genius.

User responded to Karen bragging won't get you to win madam agent.

Afterwards, Karen melded card. Next, Karen discarded card.

Afterwards, User responded to Karen good meld.

Next, Karen responded to User such encouragement to hear that!.

User melded card. Afterwards, User discarded card.

User told Karen only few cards left for you.

Afterwards, Karen responded to User yes, i just found noticed myself.

Karen discarded card.

Afterwards, Karen told User i am in this to win!.

Next, User responded to Karen well do your best.

User laid off card. Then, User discarded card.

Then, Karen discarded card.

Afterwards, User told Karen you do play very well.

Afterwards, Karen responded to User i know, i am excited.

Subsequently, User discarded card. Karen laid off card.

Karen discarded card.

Next, Karen told User yes, i am finishing my cards.

User responded to Karen not so soon.

Subsequently, User melded card.

Afterwards, User smiled.

Game ended.

Afterwards, User told Karen i finally win. good game.

User laughed. User smiled. User laughed. Next, User smiled. After which, User laughed.

Table 5.6: A sample full story generated for a rummy interaction.

(1) During Karen and User's rummy game, this happened: Karen laid off card. Then, Karen discarded card. After which, User told Karen nice.

Afterwards, User smiled. User discarded card. Afterwards, Karen laid off card. Karen discarded card. User told Karen nice again.

(2) During Karen and User's rummy game, this happened: Karen laid off card. Then, Karen discarded card. Next, Karen told User yes, i am finishing my cards. Subsequently, User responded to Karen not so soon. User melded card. User smiled. Game ended. User told Karen i finally win. good game.

Table 5.7: Two interesting spans generated by our framework from the full story in Table 5.6.

5.2.2 A User Study

I conducted a small-scale study in which 11 participants were asked to choose the more interesting story span between two options: an interesting story span generated by the framework, and a non-overlapping random span of the same story with an equal length as a baseline. Each participant was presented with an identical set of 8 questions, while the ordering of the story spans in each question was random.

5.2.3 Results

Out of 88 choices made in 8 questions by 11 participants, in 62 cases, the system-generated interesting story spans were favored by the participants, while in 26 cases, random spans were favored. Therefore, in 70% of the cases, participants identified the story spans generated by the system as more interesting than a random baseline.

Table 5.8 provides the detailed number of participants' agreements and disagreements

for each of the 8 questions in our questionnaire, along with their respective p-values from a one-tailed binomial test. I also calculated the inner-agreement among the participants, which yielded a Krippendorff's alpha [74] of 0.32.

Questions	Agreements out of 11	p-value
q1	2	0.03
q2	4	0.27
q3	10	<0.006
q4	8	0.11
q5	11	<0.001
q6	11	<0.001
q7	7	0.27
q8	9	0.03

Table 5.8: Questionnaire results for every item, including a p-value from a binomial test.

5.2.4 Discussion

The results of the study suggest that in about three-quarters of the cases, a human reader prefers a system generated-output as a more interesting short story of events, over a randomly selected span of events as a baseline. Despite the limited number of participants, these results are statistically significant in 5 of our 8 questionnaire items. This is particularly promising since the raw interaction logs used in this study had a very limited variety in terms of event types. This study demonstrates that statistical approaches can account for lack of

semantics if it is desirable to switch the domains of storytelling, and hence, limit the depth of semantic modeling for each supported domain.

Indeed, and as discussed at the beginning of this chapter, emerging use cases of computational storytelling are integrated with our lives and situated in our daily experiences, and hence, the domains over which they may be expected to tell stories can be extremely fluid. In such a situation, almost no domain semantics is desirable to assume, and thus, our approach underlines the possibility of story generation for future use cases.

Moreover, while neural story generation, which samples sentences from a model trained over large corpora of text (either in one domain or across many domains, e.g., Wikipedia), is also a promising statistical case of open story generation, it is hard to leverage neural network models to generate stories from existing past interactions and logs. A conditional sampling of such models still leads to unexpected and uncontrollable additions to the generated stories. It is possible that advances in language modeling could alleviate these problems; regardless, an evaluation metric that focuses on the perceived interestingness of generated stories is desirable, as I discussed in chapter 4. The next chapter focuses on this problem.

5.3 Chapter Takeaways

Traditional story generation methods can focus on domain semantics to derive evaluation factors: what to tell a story about, for it to be perceived as interesting. In this chapter, I discussed an approach and a framework that avoids relying on a domain's a priori known semantics, and instead, focuses on the statistical relationships between event sequences to gen-

erate what is ultimately perceived as an interesting story. I chose these statistical measures to be consistent with the theories of interest discussed in chapter 3. Using the interaction logs of a social game between an agent and human players, and without any knowledge of the game or play actions, this framework generated stories that are perceived as more interesting than a random baseline in the reported evaluation.

One of the main contributions of this framework is to showcase an open-box mechanism in which attention to perception and theories of interest can be used to generate interesting stories. I believe that the same attention can be added to some of the other approaches to open story generation as well.

Chapter 6

Telling Them Joyfully¹

In this chapter, I will attempt to build a link between the cognitive processes involved in the perception of stories and the generative process and system. This link takes the form of a quantitative measure that is informed by the theories of story interestingness discussed in chapter 3. This quantitative measure can be used in story generation approaches, especially open story generation, as discussed in chapter 4, including the approach that I introduced in chapter 5.

¹Based on:

1) **Behrooz, M.**, Robertson, J., & Jhala, A. (2019). Story Quality as a Matter of Perception: Using Word Embeddings to Estimate Cognitive Interest. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (Vol. 15, No. 1, pp. 3-9) (**Best Student Paper Award**).

2) **Behrooz, M.**, Robertson, J., & Jhala, A. (2019). Investigating the Use of Word Embeddings to Estimate Cognitive Interest in Stories. In *Cognitive Science Society (CogSci)*.

6.1 Cognitive Interest as a Proxy Measure

As I previously discussed in chapter 4, having automated evaluation metrics for generated stories is useful in general, and particularly useful for open story generation in which domain semantics are not known. In the same chapter, I explained that expanding the evaluation metrics to include a measure of interestingness is a distinctly different task than evaluating many of the surface features and the use of language. In fact, at times, a bad surface quality could *contribute* to cognitive interest. For instance, a story that adheres less than perfectly to the known sequence of verbs (thus obtains a mediocre value in terms of temporal ordering in [137], see chapter 4), may contain an *unexpected* event that contributes to cognitive interest (see chapter 3). However, too much deviation from known sequences of verbs is probably not very interesting either. This emergent balance is reminiscent of Kintsch’s idea of cognitive interest as an inverted-U-shaped function of knowledge and uncertainty [93]. To this end, I believed that the addition of proxy measures of cognitive interest would be useful to the automated evaluation of generated stories.

6.1.1 Quantitative Estimation of Predictive Inference

I focused on predictive inference which is known to be the main cause of cognitive interest according to the more recent theories, which are also compatible with the previous theories (see chapter 3, [37]). To this end, I sought to develop a proxy measure that can estimate the generation of predictive inference in the listener’s² mind. The best-known authorship skill

²Please note that by using the word “listener”, I mean a general audience, and do not mean to imply a necessarily spoken form of stories.

that can generate predictive inference is *foreshadowing*. Indeed, not all cases of foreshadowing lead to the generation of predictive inference, as the hint provided in the story can be too subtle to drive predictive inference and a hint’s connection to future events may be only revealed later in a process of postdictability. However, most cases of foreshadowing stand out to the listener as a curious case and drive predictive inference; a state of mind that is intentionally caused by the storyteller, as discussed in chapter 2.

In [10], focusing on surprise as a driving factor, a planning-based framework for generating flashback and foreshadowing is provided. While this research was an inspiration for my work, I sought an approach that can estimate the presence of foreshadowing without relying on explicit a priori knowledge, as explained in previous chapters.

6.1.2 Word Embeddings

One of the products of the advancements of deep learning is the dramatic increase in the quality of word embeddings: high-dimensional vectorized representations of words largely based on co-occurrence in large corpora. This increase in quality has even opened doors to performing analogical reasoning using word vectors [66]. Word2Vec [67], GloVe [131] and FastText [85] are three successful models for creating word embeddings; however, they are context-independent and associate a certain word in a corpus with a single vector regardless of the sentences and contexts it appears in. Newer models, such as ELMo [133] and BERT [47], take the context into consideration and associate the same word with different vectors based on the context it appears in (e.g., adjacent words, the containing sentences or story event).

Estimating the presence of foreshadowing without a formalized event sequence that

depends on a semantic model of the domain is a complicated task. Foreshadowing can take many different shapes, be causal or non-causal, or depend on domain-specific clues. However, many cases of foreshadowing involve the usage of words that co-occur in many other contexts. Thus, such words are likely to be represented by nearby vectors in the embedding space, especially one that considers the context. This is the main intuition behind the approach that I will introduce in this chapter.

6.1.3 Method

Given a short story, my method first removes all of the stop words and named-entities in it. Then, using BERT [47] embeddings pre-trained on a books corpus [184] (with 1024 dimensions), it extracts word vectors for every remaining word in the story. As previously mentioned, this model yields different vectors for each occurrence of a word in the story.

In order to simulate the linear nature of the perception of the narrative, my method incorporates a concept that I call “moving cosine similarity”. Starting from the second sentence of the story (word location b), the method calculates the cosine similarity of every word vector with the average of all of the word vectors that precede it in the story. In other words, for every word w_i , starting from the second sentence, it calculates:

$$sim(w_i) = cosine(mean([w_b..w_{i-1}]), w_i)$$

Consider the example short story seen in Table 6.1 which contains a case of foreshadowing. Calculating the moving cosine similarities for all of the words starting from the second sentence will yield a sequence of values. A chart of these values for the example story is seen

in Fig. 6.1. Assuming that every foreshadowing will consist of two main parts in two sentences (e.g., a “hint” and a “twist”), one can notice in the example story in Table 6.1 that the words *distracted* and *tired* are key words of the *hint* sentence (where the waiter is distracted), and the word *wrong* is the key word of the *twist* sentence (where the food is wrong). These words show an anomalously low amount of moving cosine similarity.

Table 6.1: An example story that contains a case of foreshadowing. The words in bold correspond to the dips in the chart seen in Fig. 6.1.

Sam and Judy went out for dinner at their favorite restaurant.

While driving to the restaurant, Judy’s favorite song played on the radio.

Sam found a parking space at the very front of the restaurant.

Sam and Judy were seated immediately and ordered their favorite food to the waiter.

He looked **distracted** and **tired** but was polite while taking their order.

Sam’s favorite song played on the radio while they waited for their food.

When the waiter returned with their food it was all **wrong!**

The waiter **apologized** and returned a few minutes later with the correct order.

Sam and Judy enjoyed their meal. They paid their tab, left a tip for the waiter, and drove back home.

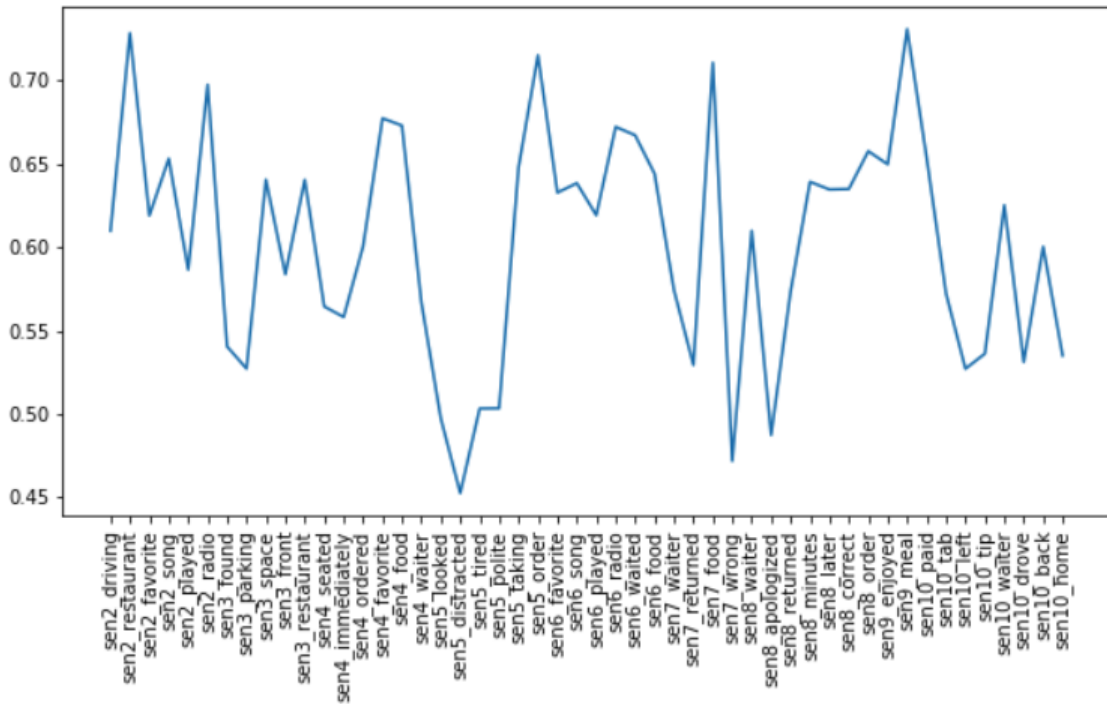


Figure 6.1: Moving cosine similarity chart of the sample story seen in Table 6.1.

To algorithmically find the anomalously low values in the sequence of values of moving cosine similarities, I wrote a simple outlier detection algorithm. Such algorithms usually have a threshold with which they detect outlier values. The algorithm seeks to find out if there are exactly two different sentences in the story where such anomaly occurs (such as in Fig. 6.1) with a fixed threshold. To this end, the algorithm gradually reduces this threshold until (and if) it finds a set of outlier words that belong to exactly two sentences in the story.

My presented method then attempts to calculate a quantitative metric “ M ” that reflects the level of anomalousness of the words involved in a possible case of foreshadowing. This measure will then be an estimation of how much predictive inference one can assume a case

of foreshadowing causes. If the method does not find any such 2 anomalous sentences, then $M = 0$. Otherwise, it first calculates A and B as follows.

- A is the cosine distance between the mean of all of the word vectors in the story, and the mean of the set of outlier words:

$$A = 1 - \text{cosine}(\text{mean}_{\text{all_words}}, \text{mean}_{\text{outliers}})$$

- B is the mean of the cosine similarity of each word in the outlier set with the mean of all others in that set:

$$\text{mean}(\text{cosine}(w_i, \text{mean}(\{\text{outliers}\} - w_i))$$

for every w_i in outliers.

A represents a measure of how anomalous the hint and twist in the foreshadowing are, by calculating the cosine distance between their means. B represents how contextually and semantically related the sets of outlier words (separated into two sentences) can be considered to be.

In order to yield M as a singular measure, the method sums A with the absolute value of B . The reason for taking the absolute value for B is that one would want to consider the semantic relationship and not necessarily similarity. If two sets of outlier words across two sentences have a cosine similarity of -1 (semantically opposite), that is potentially just as valuable for providing a hint as $B = 1$. In order to limit M to $[0, 1]$, I will note that the range of A is $[0, 2]$ and B 's is $[0, 1]$. Thus:

$$M = \frac{A + abs(B)}{3}$$

It is important to note that M is not a probability, and hence, a value of 1 (while highly unlikely) would not mean anything special. For instance, for the example story in Table 6.1, which contains a clear case of foreshadowing, these values are $A = .35$, $B = .17$ and $M = .40$.

My presented proxy measure of predictive inference and foreshadowing has a two-fold output. Firstly, it has an output of being zero or non-zero. This binary output shape is driven by the nature of foreshadowing, which necessarily consists of two places in the story that have semantic links between them. Secondly, once the method does find a pair of candidate sentences for foreshadowing, it then estimates how much predictive inference it may cause. Many forms of foreshadowing involve subtle hints that do not necessarily cause outlier word vectors. Predictive inference, however, is caused when the listener notices a form of discrepancy, inconsistency or curious detail in her perception of the sequence of events. Based on these intuitions, it is plausible to imagine that the kinds of foreshadowing cases that are capable of driving predictive inference are also more likely to involve outlier word vectors.

6.2 Evaluation

To evaluate this method and approach, I conducted a two-phased user study.

6.2.1 Study Phase I

In the first phase, I used 3 short and simple stories about going to a restaurant, going on a plane flight and a bank robbery (“restaurant”, “flight”, and “bank robbery” stories accordingly). These stories were extracted from [107] as largely mundane event sequences that lacked specificities. Each story contained 10 to 14 short sentences and all 3 stories yielded an $M = 0$ with my presented method.

I then recruited 40 participants on Mechanical Turk and asked every participant to add a “HINT” and a “TWIST” to each of the 3 stories. Participants were asked to specify the locations in the story where their HINT and TWIST would be added (between any two sentences or after the last one) and were given an open-ended text field to write their additions. While no length limit was enforced, participants were encouraged to limit their HINT and TWIST additions to 1 sentence each. They were not able to change the existing sentences in the stories. This evaluation was intended to find how reliable the proxy measure is in finding cases of foreshadowing.

Results

After cleaning the data (removing 4 participants’ data who entered random words), this step resulted in a dataset of 108 stories with foreshadowing (36 for each story). I ran the presented method on all of these stories to find out the percentage of them for which the proxy measure yields a value of $M > 0$. Table 6.2 shows this ratio, as well as the mean M values (calculated only on non-zero M values), for the augmented stories made from each of the 3

original stories, and overall.

Table 6.2: Phase 1 results. Ratio denotes the percentage of the stories with foreshadowing for which the presented proxy measure results in an $M > 0$.

Story	Ratio	<i>mean(M)</i>
restaurant	78%	.34
flight	75%	.30
bank robbery	94%	.29
Overall	83%	.31

6.2.2 Study Phase II

In the second phase, I investigated the links between M and the perceived interestingness of stories. For each of the 3 stories, I picked 2 random instances from the output of the first study: one with a high M value (randomly selected from the top 5) and one with a non-zero low M value (randomly selected from the bottom 5, excluding the ones with $M = 0$). I did not choose the stories with $M = 0$ since those are clearly missed by the proxy measure, and hence might or might not drive a high level of cognitive interest. This evaluation sought to investigate the differences in the human perception of the stories with high and low M values.

I recruited 52 participants from Mechanical Turk (different than the participants of phase I), and in a within-subject design, asked them to rate the interestingness of the 6 stories selected above on a Likert scale (1-5) and in a randomized order. This resulted in 52 ratings for each of the 6 selected stories.

Results

Table 6.3 shows the mean and median Likert rating of each of the 6 selected stories. I used a one-tailed Wilcoxon Signed-Rank test to look for statistically significant differences between the ratings of the two versions of each story (High-M and Low-M).

Table 6.3: Phase II results. The p-value is from a one-tailed Wilcoxon Signed-Rank test.

Story	Mean	Median	p-value
restaurant, High-M	2.7	3	.032
restaurant, Low-M	2.3	2	
flight, High-M	2.7	3	.085
flight, Low-M	2.9	3	
bank robbery, High-M	3.81	4	.038
bank robbery, Low-M	3.58	4	

6.2.3 Discussions

The first phase's results, shown in Table 6.2, indicate that my presented method and proxy measure perform well with a rate of $> 75\%$, across the three different original stories and cases of foreshadowing authored by 36 participants. I did not find a major concentration on a sub-group of participants for the stories with undetected foreshadowing. However, out of 19 such undetected cases, 8 of them belonged to 4 users (2 cases each). This observation can speak to the impact of individual style of writing in foreshadowing or the level of subtlety of the hints. Foreshadowing can involve long causal chains or contextual semantic links that do not depend

on words that co-occur in other contexts (and hence their word vectors do not yield high cosine similarities if trained on general corpora).

As previously mentioned, the two-fold output of my presented proxy measure also allows us to estimate how much predictive inference a case of foreshadowing makes. The generation of predictive inference is likely affected by many factors, including the more subjective experiential interests discussed in chapter 3. Thus, this proxy measure's estimation is mainly based on the intuition that if a set of outlier words are semantically farther away from what the rest of the story has been about, they are more likely to raise a question mark for the listener and drive predictive inference. In simple terms, the farther such distance is, the bigger the mental question mark of the listener can be. For all 3 of the original stories in the evaluation, the average M listed in Table 6.3 is about .30, with a maximum value of .42, .40, .37, minimum of .25, .18, .18, and a standard deviation of .04, .06 and .06, respectively. These results indicate that the proxy measure has some level of variation, but the variance is small enough that one can categorize the level of estimated predictive inference in "high" and "low" groups. It is plausible that for other datasets or longer stories this variance could grow.

Consistent with such categorization of "high" and "low", the second phase of the study found statistically significant differences in the perceived interestingness of randomly chosen stories with high and low M amounts for 2 of the 3 original stories (restaurant and bank robbery). The samples of the other story (flight) with high and low M amounts did not show a statistical significance in the difference of their perceived interestingness by the participants.

It is noteworthy that the bank robbery story shows higher perceived interest levels than the other two stories, as a plot that involves *danger*, one of the Instinctive Interests (or

“absolute interests” [152]); see chapter 3. This point also introduces a connection to the results of the experiments discussed in that chapter; as interests interact and the effects of experiential interests can make the measurement of variations in perceived cognitive interest to be much harder to detect.

6.3 Chapter Takeaways

As I explained in chapter 2, the telling of interesting stories involves an understanding of the perception of them by the listener. In chapter 3, I first explained my work on understanding the qualities of the perception of stories and what causes interest in them and then reported on my expanding of existing theories and an evaluation. Then, in chapter 4, I discussed the generation of stories, the role of reliance on semantic models, the importance of evaluation and the current state of automated evaluation metrics and proxy measures of story quality. In chapter 5, I outlined a new approach to open story generation with a focus on previous events and interactions and reported on an evaluation of it.

With that perspective in mind, in this chapter, and with a goal of informing the story generation process of the cognitive processes involved in the perception of stories, I outlined my approach for deriving a quantitative proxy measure of predictive inference using word embeddings. Through a user study, I offered reasons on why this proxy measure largely correlates with human judgment of story interestingness. This quantitative proxy measure could be used in a variety of story generation approaches, particularly useful for the ones not relying on domain semantics (open story generation); e.g., the approach presented in chapter 5, or controllable

neural story generation approaches such as in [169].

Part III

Stories in the Wild

Chapter 7

Musical Stories¹

In this chapter, I report on using story generation techniques in the music listening context. During an internship at Spotify, the popular music streaming service and company, I created a prototype that uses story generation techniques to augment the songs with background and contextual information. This prototype adds a snippet in between every two songs which states a property of the second song or a semantic link between the first and the second song. I call this snippet a “segue”. The overall music listening experience enabled by this prototype, therefore, takes the form of a sequence of songs and segues that can augment the music listening experience with background information (e.g., on a smart speaker) and provide a story-like music-and-segue narrative. Segues have the opportunity to cause predictive inference in the listener’s mind, as they attract the listener’s attention to the next song and what it may be. Story interestingness and predictive inference are discussed in chapter 3.

¹Based on: **Behrooz, M.**, Mennicken, S., Thom, J., Kumar, R., & Cramer, H. (2019). Augmenting Music Listening Experiences on Voice Assistants. In *International Society for Music Information Retrieval (ISMIR)*.

7.1 Motivation

Voice-enabled devices, such as “smart speakers” like Amazon’s Echo, Apple’s HomePod, Google Home, or Sonos One, have reached the mainstream. In particular, listening to music is a popular use case for such devices [124, 155]. Finding music to listen to and discovering music on these devices can be a challenge as the interactions supported by voice-enabled speakers are relatively limited by current interaction models.

But as music is deeply rooted in culture and group identities [77, 87, 146, 180], music fans are often also interested in contextual information in addition to the audio content consumption. This deep root in culture is reminiscent of the prevalence of stories in human social interactions and communication (see chapter 2).

Prior research suggests listeners employ music search to learn about and explore new content to consume. Listeners seek background information to stay informed about their favorite artists, genres, and songs and use it as a relationship builder with others [108]. This exploratory mindset, however, is relatively rare on music streaming apps because catalog-based entity search does not support this user need well [79].

Augmenting listening experiences and conversational interactions have the potential to support these exploratory user goals but leveraging them for a good user experience remains a challenge. Learning about background information is sometimes a part of the listening experience itself. Often, such information is presented together with the music playback to contextualize the content.

For example, user interfaces of several music streaming services, such as Apple Mu-

sic, Pandora, and Spotify, include a section for additional information beyond basic track metadata for artists, albums, and playlists. Sometimes, songs are contextualized further by displaying the lyrics, stories, or background information associated with certain parts of the songs (e.g., “Behind the Lyrics” feature on Spotify [179]).

In the following sections, I outline my method for using story generation to automatically augment the voice-based music consumption with background and contextual information, explain my prototype system, and share the results of a qualitative study using this prototype.

I report on making the following contributions in this work:

- Introduction of a type of content augmentation to contextualize voice-based content consumption with background information in Section 7.3.
- Detailed design of an approach taking playlists as input and utilizing weighted graphs to generate textual music augmentations, inspired by story generation in Section 7.3.
- Identification of best practices for using augmentation and conversation in voice-based music consumption in Section 7.5.

7.2 Related Work

7.2.1 Listener Information Needs and Music Search

When listeners search for music, they have multiple information needs that they may be trying to fulfill. These user needs help to shape how listeners approach their music search goals. For instance, listeners may be in the mindset of looking for something specific or they may be in the mindset where they are open to multiple types of music-related information. Prior

research has suggested that users of a streaming music service have distinct mindsets when they are searching for music [79]. In a focused mindset, users have one particular item in mind. Catalog, entity-based search interfaces favor this particular mindset and queries that align with the structure of available metadata. In an analysis of Google Answers queries, Bainbridge et al. [11] found that users typically (81.3% of the time) expressed needs through bibliographic queries, using performer, the title of work, or date of recording. Li et al. [106] also observed that typed searches on a streaming music platform are typically focused, suggesting that the modality and design of the current feature supported this type of mindset.

Listeners also have broader information needs that are not met by catalog-based entity searches commonly supported in online music services. Lee et al. [101] observed that people use cloud music services that store listeners' music libraries to listen to music that they were unfamiliar with, suggesting that music discovery and exploration is an important user need. In addition, listeners indicated they search for information about the artists and music for learning purposes [100]. Users of a streaming music platform, however, tended not to use the search feature to deeply learn about a specific type of music and left the platform to fulfill that need [79].

7.2.2 Voice Assistants and Music Consumption

Voice-enabled speakers currently allow music listeners to search for content (e.g., by saying “play Jazz” or “play Time by Pink Floyd”) and control the music playback (e.g., play/pause/skip and volume controls). In fact, these basic playback controls form the most common category of user commands [155]. While many of these speakers can be used in con-

junction with a secondary device that has a graphical user interface (GUI), voice interaction is increasingly becoming a primary modality for consuming music [155], which increases the importance of evolving and improving the music experience through voice. In [124], for instance, it is reported that 34% of the respondents said that the time they spend on music consumption via such speakers is replacing the analogous time spent on smartphones.

Notably, though, the voice-only smart speaker experience does not offer much in the way of discovery or background information and such lack of contextualization and grounding can reduce music discovery and listener's emotional investment [166]. An article in [24] frames this issue as a lack of metadata in the music content delivery through voice and argues that if these problems are addressed inefficiently, it could disproportionately hurt less-known artists who are more dependent on discovery on the part of listeners. What could serve as a solution to this issue is utilizing the voice interface to do more than "command-and-fetch" for music and allow the voice output to provide more information. Additionally, a conversational experience could enable a dialogue where, for instance, a music-and-spoken-word journey can follow a listener's request to learn about the early years of jazz or the founders of hip hop (e.g., as discussed in [5]).

My work focuses on contextualizing the voice-based music experience with relevant background information. This idea shares similarities to music radio shows, where the hosts provide relevant information about the content they play and add other talking points in between songs. In [18], radio's interaction of speech and content is framed as a special kind of narrative, in which the DJ or radio host is the narrator. One of the main challenges in creating an experience like radio shows is maintaining the "flow" of the music, and balancing the

spoken words and songs, as this is one of the main skills of the radio hosts [6]. My user study seeks to learn more about how to achieve a balance between this flow and providing background information.

A related use case of conversational agents is providing information in support of an ongoing activity or experience. Such experience-driven information is not necessarily an “augmentation”, but shares a similar goal of enriching the original experience and meeting a contextual user need. In [95], for instance, authors introduce an agent that can guide people in a public computer museum during their visit.

7.2.3 Using Story Generation

Story generation is the problem of automatically selecting a sequence of events that meet certain criteria and can be narrated as a story [107]. Story generation and my approach to augment the music listening experience share the goal to generate sequences of textual content given specific constraints.

While there are many different approaches to generate stories [49, 113, 183], ours is similar to planning-based approaches which also commonly use graph representations to map the space of story events and the possible constraints of a valid or optimal progression of the storyline. In [107] such constraints are reflected by logical precedence rules, while my method utilizes edge weights and pathfinding to extract a preferred storyline. Similar to PlotShot [38], I apply a graph-based approach to generate a sequence of text for a given form of input media. Inspired by these different approaches, this approach takes playlists as input and utilizes graphs

with edge weights that reflect content preferences to generate textual music augmentations.

As I discussed in previous chapters, many approaches have been taken to story generation, including case-based reasoning [49], planning [183] and more recently, deep learning [113].

Many of these approaches involve using a graph in which the space of story events and the constraints of the progressions of the storyline is mapped. My approach also uses a graph to represent the possibility space; however, instead of commonly-used logical precedence rules (e.g., in [107]), I take an optimal path finding approach, where the existence of certain nodes in the path can make the selection of other nodes more or less likely. In [38], a graph is used to represent a planning-based approach to story generation, in building a system called PlotShot which generates textual stories from a sequence of photos. Although the authors do not explicitly frame their work as an augmentation task, the PlotShot system can be viewed as an augmentation tool to add a narrative to a set of photos. Regardless of augmentation, this work is similar to the PlotShot system in that it uses similar story generation techniques to generate a sequence of text for a given form of media.

The augmentation material that this approach generates is initially textual. But given the presentation of the output via synthesized speech along with a set of songs, this work is also related to the problem of narrative generation in multimedia research. For instance, in [185] authors describe a system for creating personalized videos that recount an event, using multiple source video clips. Similar ideas have been recently incorporated into new features for smartphones, where a movie is automatically generated from a user's recorded photos and

videos from an event [7, 173]. As another example, the system in [27] focuses on creating matter-of-opinion video documentaries from recorded interviews on a given subject.

7.2.4 Voice and Interactive Narrative

Interactive narrative is a form of narrative experience where the user creates or influences the progression of the unfolding drama. The main goal in this field is user immersion, where the user believes that they are an integral part of the experience [141]. Although interactive narrative has been used for education and training [147], the most common use case of it is gaming [114, 143]. An interactive narrative can be experienced through various mediums, such as visual, text-based, or voice interfaces. Conversational user interfaces introduce an opportunity for interactive drama experiences. For example, BBC recently released *Inspection Chamber* [16], an experience in which a pair of voice characters interact with the user to find out what imaginary creature the user is, while her/his answers affect the plot and can change the ending. In [69], the creators of a Netflix TV show called *Stranger Things* incorporate the frequently highlighted use of walkie-talkie devices on the show and engage users in a similar way using voice input and output to have them interact with one of these characters and affect the plot progression in a given scene.

While this work of augmentation does not involve an interactive narrative experience, I used Wizard-of-Oz (WoZ) to include conversational capabilities in the evaluation in order to learn about the effects of interactivity on voice-based augmentation of music content.

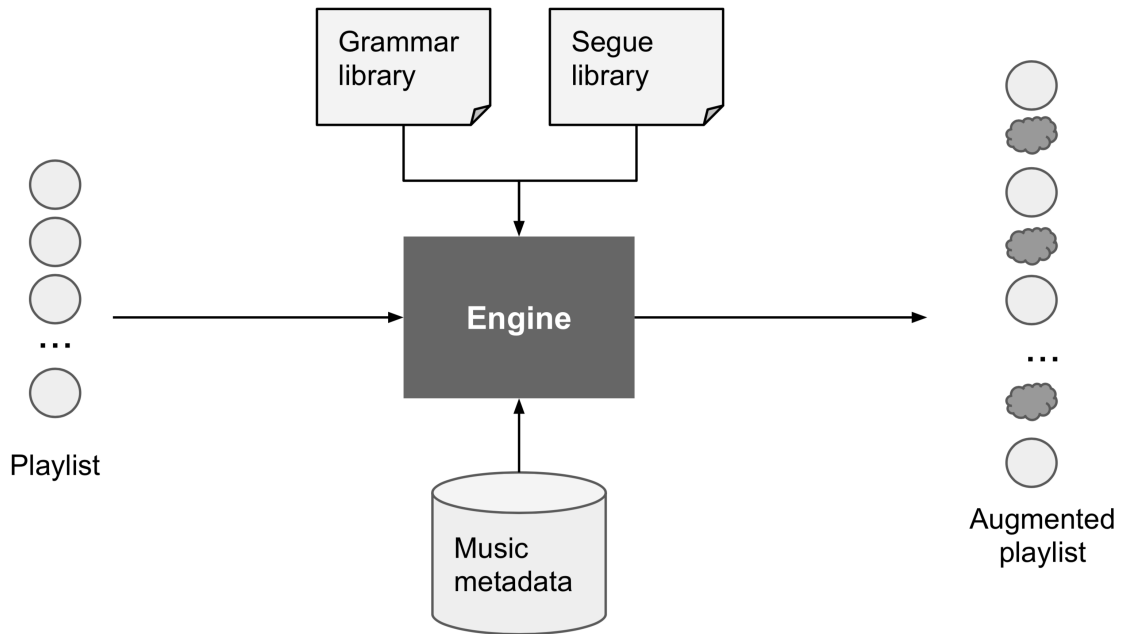


Figure 7.1: The system architecture of the prototype.

7.3 Approach and Prototype

The prototype takes ordered playlists as input, finds relevant background information and relationships for the songs it contains, and chooses a subset of that information for being used in the output (illustrated in Fig. 7.1). I call every piece of information that comes in between two consecutive songs a *segue*. Every segue describes a predefined property, such as some information or characteristic, of the next song or a relationship between the current and the next song.

While this approach is not limited to playlists as input, I decided to use them as a starting point given that songs in a playlist typically contain more variety in the metadata as opposed to an individual artist's album. Moreover, songs in a playlist often have an implicit

Table 7.1: Songs' metadata examples.

Property	Applicable Entities	Example
Name	Artist, Album, Song	Drake, Scorpion, Wild Thoughts
Musical origin	Artist	Los Angeles
Genre	Album, Song	Hip Hop
Mood	Song	Calming
Qualitative fact	Artist, Song	Rihanna's real name is Robyn Rihanna Fenty.

reason for having been grouped together (e.g., being of the same genre, suiting a specific mood or situation [121], artist similarity, etc.)

I kept the original order of playlist songs to preserve possible semantic reasons behind curation by playlist creators.

Music Metadata

My prototype uses a set of metadata and background information about songs, artists, and albums. Table 7.1 shows some sample entries of qualitative facts about songs and artists consisting of short extractions from publicly available sources of background information (e.g., Wikipedia).

Table 7.2: Examples for segues, their logic description, and samples for their realized text.

Segue Type	Logic Description	NLG template	Realized text
NullSegue	Always a match regardless of the songs.	N/A	N/A
MundaneSegue	Always a match regardless of the songs.	Next song is next_song_name by next_song_artist_name.	Next song is Time by Pink Floyd.
ArtistOriginJump	Musical origin of the previous song's artist is different than the next one's.	From prev_city where prev_artist_name's musical origins are, to next_city where next_artist_name's are.	From Los Angeles where Tupac's musical origins are, to New York City, where Biggie's are.
SameYearSameArtist	Previous and next song share the same artist and release year.	Just like the last song, the next song is from next_song_release_year by next_song_artist_name.	Just like the last song, the next song is from 2007 by Ri- hanna.

Segue Library and Grammar Library

Table 7.2 contains a few examples for segues. Each segue has a natural language generation (NLG) template resulting in a “segue text” when realized. The segue library contains 21 segues. The authoring effort for creating new segues simply depends on the complexity of the segue logic.

Inspired by the story generation concept of grammars [30], I defined a simple construct in the system to allow prioritizing authored sequences of segues that are presumed to be interesting. For instance, by preferring a sequence of `ArtistFact`, `ArtistOriginJump`, and `ArtistFact`, an augmentation can focus on the background of songs and their artists. Grammars are an instrument for professional authorship and editorial opinion to be reflected in the system.

7.3.1 Generating a Sequence of Segues

First, this prototype accesses available metadata about songs, artists, and albums appearing in the playlist. Then it finds all the matching segues for every two consecutive songs which results in a list of segue options for each such position. For the entire playlist, there would be a list of these lists, which I call the *story possibility space*. Given that the choice of a segue at each position in this space is independent of other positions, the story possibility space forms a graph and the search problem for finding a sequence of augmentations becomes a problem of finding the best path in this graph. To do so, I use a set of heuristics and preferences which are reflected in a *weighting function*.

These scores are assigned as weights to the edges that represent those transitions in the graph.

$$\begin{aligned} \text{weight}(s_1, s_2) &= \text{diff}(s_1, s_2) \\ &+ s_1^{\text{pref}} + s_2^{\text{pref}} \\ &- \text{lengthiness} + \text{silence_reward} \\ &+ \text{playlist_reward} \\ &+ \text{positional_preference} \end{aligned}$$

Several variables enable weighting absolute and relative preferences. $\text{diff}(s_1, s_2)$ enables avoiding repetition between consecutive segues. Static “segue preference scores” s_i^{pref} give specific segues authored preference. For example, pointing out a change of genre between two consecutive songs might be more interesting than simply stating the title and artist of the next song. Terse responses are often preferred in conversational interactions [40], hence *lengthiness* punishes a segue if it has a long text and *silence_reward* rewards a graph edge if the previous segue is long but the next segue is `NullSegue`. *playlist_reward* represents that some segues fit better to a specific type of playlist, such as `ArtistQualFact` in artist-focused playlists. *positional_preference* is used for segues that make only sense at a specific part of a playlist. For example, a playlist introduction with a short authored description only makes sense at the beginning.

Given a weighted graph, I first look for and choose any possible grammar matches. A grammar is a match if there exists a path in a sub-graph of the story possibility space, where the

sequence of nodes in that path matches the grammar's sequence of segue types. Edge weights do not have a role in finding a grammar match. If two grammars overlap, I choose the path representing one of them at random.

For the portions of the story possibility space where no grammar match is found, I use the edge weights to find the best path, one with the heaviest sum of weights. If a given portion of the overall graph that needs pathfinding is larger than 5 playlist positions, I find the path step by step in windows of size 5. In doing so, I ensure that each such window does not contain any segue types that exist in the previous window, and hence avoiding local repetition of segue types.

To exemplify conversational interactions, I identify possible interaction points in which one could trigger a short dialog and let the user response determine which segue option comes next. I do so by checking against simple logic definitions, e.g., if there are specific types of segues in the next list of segue options (see Table 7.3.)

After the full graph path is determined, I use the realized segue text of the segues in the chosen path and insert these segue texts in between the songs. An example excerpt of an augmented playlist is shown in Table 7.4. My prototype can generate augmentations for any given playlist as long as it has access to the metadata for the songs in that playlist. For the evaluation, I decided to focus on three popular types of playlists to start with those based on an artist, a genre, or listener popularity.

Table 7.3: Examples for conversational augmentations.






Voice Prompt	User Response	Voice Response
From when do you think this last song was?	Correct	That’s right. But the next song, called Shook Ones, Pt. II takes us into a different era. All the way to 1995. (DifferentEraSegue)
	Wrong	Actually, it’s from 2007. The next song called Shook Ones, Pt. II and [...]. (DifferentEraSegue)
Question! Are you more interested in the artist’s background or the genre?	Genre	The genre of the upcoming song is called “Latin Trap”. (NextGenreSegue)
	Artist	Next song is by Cardi B. Here’s a fun fact about their biography... (ArtistQualFact)

7.4 Evaluation

To better understand how my method of adding contextual information to smart speaker experiences affects music listening, I conducted a two-phased study within Spotify.

In phase 1, I gathered feedback from two professional writers who are familiar with the music domain to elicit expert feedback on the content of the segues. They received the written output of the prototype generated for one representative example of each playlist type: artist, genre, and listener popularity. While I invited them to provide any type of feedback, I

Table 7.4: Example excerpt of an augmented playlist.

...
—————  Juicy by The Notorious B.I.G. —————
Here's The Next Episode by Dr. Dre and Snoop Dogg.
—————  The Next Episode by Dr. Dre —————
Now switching from the 2001 (Explicit Version) album to one called The Best of 2Pac - Pt. 1:
Thug.
—————  California Love by 2Pac —————
The last song was from 2007. The next song called Shook Ones, Pt. II takes us into a different
era. All the way to 1995.
—————  Shook Ones, Pt. II by Mobb Deep —————
Just like the last song, this song was released in 1995.
—————  Gangsta's Paradise by Coolio —————
The last and the upcoming song both are described as dark groovy.
...

specifically asked them to share their views on the contents of individual segues and describe how they would approach writing similar content from a professional perspective as writers. After they returned their comments, I conducted a semi-structured interview with both of the writers which took about 45 minutes.

In phase 2, I conducted an internal evaluation with nine Spotify employees (four female, five male) from various parts of the organization to identify potential future improvements and establish a first understanding of user needs. Participants were in their early 20s to late 40s from non-technical functions (such as design, marketing, or operations) and located in various locations across the United States.

Each session included a semi-structured interview in which I asked participants about their previous experience with voice assistants and whether or how they look for additional content around music. Each participant was asked to listen to a demo audio file for one of the three playlist types. After answering a short questionnaire, they also interacted with my envisioned conversational experience in a short WoZ demo where an experimenter controlled which content to play. Each playlist type was presented to three users who were randomly assigned to a condition.

The demo consisted of ten shortened songs (first and last 15 seconds) and ten segues (one intro segue, nine transition segues) which were generated using my proposed method and then read by a text-to-speech (TTS) engine. Overall, they had a duration of 5:30-5:50 minutes. The short WoZ section to convey the conversational experience covered three songs only, but between the songs, the TTS voice prompted the participant with a potential question such as “*Question! Are you more interested in the artist’s background or the genre?*”. Depending

on the answer the experimenter chose the next audio file to play to continue the experience. Table 7.3 shows two examples.

I recorded and transcribed all of the sessions. Two other researchers went independently through the transcripts, first categorizing them for their relevance to the stated research questions and then doing an affinity analysis [22], moving relevant quotes between the high-level categories, to derive my findings.

7.5 Findings

I identified various factors that influence the perception and usefulness of including contextual information in music listening experiences.

7.5.1 Addressing Listener Needs and Contexts

Music is consumed in vastly different situations, playing a different role for the listener's needs in each one. I found that listeners' perceived usefulness of the voice-based augmentations heavily depends on the situation and its unique needs.

Augmentations enable music discovery and education. Augmentations are well received when listeners are in an exploratory mindset. The participants expressed a special interest in using the voice augmentations to learn about content that is new to them. P4 said: “[*Talking about a playlist containing new songs*] I’m like ‘Wait, what band is this?’ [...] ‘What other songs can I listen to from them.’ ” P6 described their interest in being able to learn about

(niche) genres through such augmentations: *“I feel like metal would work really well for this because a lot of bands have a lot of history behind them [...] it’s the opposite of trendy [...] people are still listening to music that was written and performed 20 years ago consistently.”*

Similarly, Editor 2 saw them as a way of discovering less-known artists by providing information about them: *“[When choosing music automatically] you might end up skewing the information toward [...] the top-selling artists of all time; yet obviously there are hugely influential artists that have not sold a lot of records but have impacted other artists and bands.”* Lastly, P1 brought up the need to identify the right occasions for adding information: *“I like that it’s just another way to get to know an artist that you already like and I would potentially like if it was getting to know an artist that you don’t know. What I wouldn’t like is if it’s in between.”* This highlights a potential for leveraging the listener’s level of affinity for an artist they already know, or the predicted level of affinity for a new artist, in determining the quantity or focus of the augmentations.

Activities determine needs for and appropriateness of augmentations. Music often supports a specific listener activity. I found that activities with low cognitive load, such as doing chores or cooking, were commonly mentioned as appropriate contexts for voice-based augmentations. P8 said: *“The perfect experience [is] if I’m at home doing something fun like cooking or something not fun like cleaning.”*

Activities that require a higher level of focus but that listeners consciously choose to support with background music were perceived to be less suitable. Participants mentioned several examples where the music is serving such an activity-supporting focus like working

out, studying, or relaxing and felt that any addition to the music could get into the way of that primary activity. *“I need [the music] to keep the motivation going, keep the music going.”* (P8)

7.5.2 Selection of Appropriate Content for Augmentations

The next category of my findings relates to the content of the augmentation and what it focuses on.

Personalizing the augmentations improves the experience. The level of affinity with an artist or genre varies significantly across listeners, and the same is true for the level of familiarity with background information. For example, using a sub-genre to describe a song might be very interesting to someone familiar with the general genre, but vague and uninteresting to someone who is only a casual listener of that type of content. P3 said: *“[...] a high, medium, low, [or] novice/expert setting [would be good], because I’m not an expert on this, so I don’t understand [some of the segues].”* Similarly, P1 saw an opportunity to point out to them if they are listening to an artist for the first time: *“Say it’s the first time I’ve listened to an artist, I think it would be cool to learn more about that artist.”*

Another frequently mentioned interest for personalization was to allow the listener to adjust the topics that the augmentations focus on (e.g., artist life or genre information). P9 said: *“If I could somehow customize like what’s being said by the voice to choose like facts or historical whatever, I think that’d be cool.”* Editors had similar views. Editor 2 said: *“we’ve got one end of the spectrum there is music nerds. They’ve already put their hands in the air and said, ‘Please give me more as much as you can.’”* The same editor then drew a parallel between

customization of content and augmentation: *“can I add another layer of personalization to this which is, please make [the augmentation] minimal [or] please tell me as much as you can about this artist or this genre.”*

Augmentations could explain recommendations or present relevant news.

My study subjects mentioned other types of information that would be useful for them to hear. Music listeners increasingly delegate their choice of music content to streaming services, which use various algorithmic and machine learning methods to choose songs that they believe the listener might enjoy. However, listeners usually do not get any explanation for why a particular set of content is chosen for them. P4 said: *“It kind of guides you to know how they’re piecing together this playlist for you. It’s like, ‘this is why we’re playing this song for you’,”* and P3 mentioned: *“A lot of times for [automatically generated playlists], I’m like why do I have this song, it would be great if [the voice] could tell me.”* Alluding to the same point, Editor 1 noted: *“With just the bare information the name, the title, and to give more information and background obviously [one can] provide a much deeper experience for users and give users the reason why they should continue listening.”*

The contextual needs of music listeners often extend to their awareness of the current happenings in the music world. Most prominently, participants expressed interest in hearing about tours and relevant news headlines. P1 said: *“If they were on tour in my area, that’s something I’d want to know,”* and P5 mentioned: *“There’s a lot of news always with musicians, whether it’s a controversy or other things [...] if you had some of that, like why is this song popular right now or what’s going on with this song.”* Editor 1 brought up the same point, and

discussed the following as an example: *“Let’s say [an artist] passes away [...], and you insert a little nugget of information to inform people about that. And then the next song is [by the same artist]. I mean [...] that might make it even more important for maybe someone to listen.”*

7.5.3 Appropriate Presentation of Augmentations

The last category of findings offers insights about the delivery and presentation of the augmentation content.

TTS voice needs to be trustworthy, high quality and fitting. The synthesized voice in which the segues are presented to the listener was one of the most common topics brought up by the participants; most prominently, the quality of the TTS voice as described by P5: *“With the DJ kind of idea, I think the sound of the thing makes a big difference; so [...] that computery voice takes me out of the moment”*. Despite the quality of the TTS, participants seemed to establish a connection with the agent behind the voice, and explicitly expressed a preference for knowing or at least being able to trust the agent. For instance, P2 said: *“Using someone’s voice who is an authority on the genre or playlist [is better] [...] there’s a difference between that voice telling me little tidbits and somebody like [reference to a Jazz musician].”*

Participants mentioned they would like specific properties of the voice, such as gender and accent, to be personalized, either based on the current content or their general preferences. For instance, P4 said: *“I like [it] when people have the Google or Waze, the driving apps, and you can change the accent.”* P5 noted: *“I think it would be cool if it was kind of genre-based [...] yea if it’s tied to genre or playlist type of thing.”* Editor 1 pointed out to voice’s gender as

well, saying: “*It’d be really jarring to hear like a very male voice [on] Ani DiFranco or Riot grrrl playlists or a very feminist playlist*”.

Augmentations should not be frequent.

Participants expressed a preference for segues that connect the previous and the next songs (e.g., by highlighting similarities or differences) over segues that focus solely on the next song. For example, P3 said: “*I like this [...] it tells me a little bit about what I just listened to [...] and then it sets me up into what the next song is going to be,*” and P5 mentioned: “*I [liked] that some of them attempt to link the previous song to the next song.*” While semantic continuity is valuable, the frequency of augmentations should not be too high, and segues should not come in between every two songs. I included a representation of an intentional skip (NullSegue in Table 7.2), but it formed either zero or just one out of the 10 generated segues that each participant experienced. Five of the participants (P1, P2, P4, P5, and P8) believed the augmentation was too frequent. P1, for instance, said: “*I definitely in no scenario want [to hear the segues] after every single song*”. Lee et al. [102] found that different user personas have a varying desire for engagement when interacting with music information retrieval systems, which needs to be taken into additional consideration when designing such augmented listening experiences.

Participants enjoyed the conversational augmentations.

My conversational augmentations showcased the ability to ask about the music that is being played, and this was well received by the participants. Most of them (seven out of nine) counted the conversational demo as more fun and interesting than the non-conversational case.

When probed on the reasons, participants frequently pointed out the ability to interact. P4, for instance, said: *“You kind of feel like there’s this other entity that you’re having a conversation with.”* In another example, P7 said: *“I think I like this better [than the non-conversational demo]. It was more fun [because of] the interaction aspect of it.”* However, two of the participants (P1 and P3) could not imagine themselves using the conversational experience in any situation and generally disliked it. Both participants attributed this dislike to usually preferring a “leaned-back” music consumption mindset, as P3 said: *“I don’t want it asking me questions. I actually hated it. It wasn’t lean back and was trying to get me to interact...”*

7.6 Discussions

The results indicate that augmenting voice-based music consumption with story-like background information addresses some of the listener needs that are commonly ignored in current experiences [5,24]. But similar to how different listening situations affect musical preferences [123], one would need to investigate situation-specific preferences for augmentations to understand when music listening is a passive [32], flow-like [48] experience which should not be interrupted.

It is useful to prioritize a story topic to increase narrative coherence [58]. In the music context, such narratives could be based on a variety of topics, such as recent events, genres, or artists, all of which were mentioned by the study participants as well. For instance, a dynamically generated augmentation about the history of a genre could focus on songs that represent the turning points of it or have other musical significance. Access to more metadata

and large semantic models that capture music-related relationships between various entities can help a story generator in achieving this goal.

In terms of presentation, my evaluation suggests that the quality of the TTS engine seems to be particularly important for music listening experiences. I suspect that the imperfections of the TTS might be more apparent due to a general focus on the audio quality, both for music and voice output. In other use cases for voice assistants, the focus is often more on retrieving the requested information; however, this hypothesis requires further research.

Changing the voice's accent or gender, based on explicit listener preference, was stated as an interest by several participants; doing so automatically, such as based on a listener model or audio content, is not only very difficult but also poses the risk of reinforcing stereotypes of societal and cultural associations for certain types of music.

To minimize the negative effects of breaking the audio flow of the music experience, a smoother transition between augmentations and music content is needed. For example, by matching audio properties of the augmentation with those of the surrounding music content, similarly to the techniques used by radio show hosts to match the nearby songs in their ending and beginning [6].

7.7 Chapter Takeaways

The music listening domain introduces an opportunity for story generation to be presented to millions of people. The combination of the interest in music, the background information and news around music, and the general appeal of the stories offer a promising outlook.

Through my internship with Spotify, a popular music streaming service and company, I created the presented prototype that uses storytelling to augment sequences of songs with background and contextual information about the songs. Moreover, certain properties of this augmentation and story-like experience can evoke predictive inference and consequently cognitive interest in the listener's mind. I also reported on an evaluation, including semi-structured qualitative interviews that addressed my goal of understanding this new paradigm of music experience.

Chapter 8

Hedonic Shopping Stories

In this chapter, I report on using story generation techniques in the domain of online shopping, made possible through an internship at eBay, the popular online shopping platform. I will explain a prototype system that focuses on topic interests of experiential interests, as explained in chapter 3. The prototype generates short story-like text snippets about a category of products that a user has an interest in. This text snippet can then be used in a voice-enabled device, such as a smart speaker, to update the user about said product category via voice. I will then report on a qualitative user study aimed at evaluating this approach and paradigm of experience.

8.1 Motivation and Related Work

Online retail is a thriving market. In 2017, e-commerce sales amounted to 2.3 trillion USD worldwide, and they are projected to be over twice that amount in 2021 [165]. Online

shopping is likewise a commonplace user behavior. An estimated 1.6 billion users around the world purchased goods online in 2017 [163], and 77% of Internet users in the U.S. (representing 67% of the population), purchased products online in 2016 [164].

In such a large and growing market, recognizing various shopping motivations is crucial to both sales growth and consumer satisfaction. These motivations are usually recognized in terms of being either “goal-oriented” and “utilitarian”, or “recreational” and “hedonic” [73,91]. Utilitarian shopping is task-oriented and is based on efficiency and rationality, while hedonic shopping is more experiential and is based on curiosity and pleasure [150].

Multiple factors such as convenience, time efficiency, availability of aggregate information (e.g., product reviews), and the interactive nature of the involved interfaces have highlighted the utilitarianism of the online shopping experience [42]; however, hedonism is not exclusive to the brick and mortar retail and is reported to be present in online shopping as well [42, 130, 151, 175].

8.1.1 Hedonic Shopping Motivations

Hedonic shopping motivations were first introduced as “non-functional” motivations [170], and have been regarded as experiential and emotional [42, 151]. The existence of distinct hedonic motivations of shopping has been confirmed in prior research [175, 176, 181]. In [9], for instance, six such motivations are introduced as follows: *Adventure shopping* refers to shopping for stimulation and “being in another world”. *Social shopping* is the enjoyment of shopping when it is done with family and friends. *Gratification shopping* is about the stress-relief resulted from shopping. *Role shopping* is buying items for others; e.g., as gifts. *Idea shopping* refers

to keeping up with trends, new fashions, and innovations; and finally, *value shopping* is about shopping for sales and finding bargains.

Even more hedonic shopping motivations have been introduced in prior research. Some of them such as *status* [181], which refers to the sense of superiority in receiving service from others, are less applicable to the current forms of online shopping, while other motivations such as online *privacy* are exclusive to online shopping [176]. As far as thirteen distinct hedonic motivations of online shopping are found in a focus group study in [176], and the ones that are most frequently discussed by the participants in that study are *bargaining* (value shopping), *privacy*, *social*, and *learning trends* (idea shopping).

While some hedonic motivations, such as *value* or *gratification*, are automatically achieved as an inherent part of the act of shopping, others might need or benefit from more explicit facilitation to meet user's needs. For instance, *role* shopping can happen without any explicit support from an online shopping platform, but it can also benefit from a "gift finding guide". Similarly, a need for *adventure* in shopping can be met with extensive web browsing sessions, or can be supported explicitly by novel applications such as a Virtual Reality Mall [103]. As yet another example, online services can facilitate the *social* aspects of shopping by providing information about what one's friends like and buy. The popular social network Instagram has recently introduced features that move in this direction [171].

In this chapter, I report on my focus on *idea shopping* (or *learning trends*). This hedonic motivation is highlighted by many researchers [9, 170, 176] and is grounded in McGuire's categorization theories [117], which explain the human need for structure, order, and knowledge, as well as Festinger's objectification theories [57], which explain one's need for external

information in order to make sense of herself. In the absence of explicit facilitation for this hedonic motivation, some users browse online shopping websites for long stretches of time to learn about the current trends while there is not necessarily a goal of making a purchase [25]. This behavior is also linked to experiencing positive affect [26]. The work reported in this chapter focuses on more explicit facilitation of this hedonic motivation through voice interfaces, and it would be interesting to see if similar effects are present in the absence of visual stimuli.

A connection between the hedonic motivations of shopping and experiential interests in the perception of stories is made when a user seeks information about shopping for pleasure (i.e., hedonic shopping motivations) and receives a story-like experience about a product category that she likes (i.e., experiential interests in stories).

8.1.2 Voice Interfaces and Shopping

As mentioned in chapter 7, the popularity of voice-enabled devices has recently seen a significant surge. According to research conducted by eMarketer [53], the adoption rate of “smart speakers” is only second to the emergence of smartphones in recent years of technology trends and the usage of these devices has surpassed that of wearables [81].

According to a study in [155], user interactions with smart speakers are most frequently related to music, smart home controls, and general knowledge (e.g., the weather). Purchasing, however, is reported to only form 0.3% of user commands. According to another study reported in [122], only 7% of smart speaker owners have *ever* used these devices to make a purchase. Amazon’s Alexa voice assistant, employed in the “Echo” smart speakers, is reportedly [172] the most widely used voice assistant in the market, but it may not be actively used

for shopping [174]. While there are predictions that suggest more users will make purchases through voice interfaces in the future [43], given the current trends discussed above, it is not clear if such forecasts are taking place.

I believe that the conversational nature and the human likeness of the voice interface make it a more appropriate interface for a focus on the hedonic motivations of shopping. In the following sections, I will describe a simple generator that provides users with sales trends and background product information for a given category of products. To this end, I needed to investigate the benefits and limitations of providing this hedonic value to the users through a voice medium. Moreover, I wanted to investigate how users may interact with such experience in the context of their daily lives and activities and what its possible effects on their shopping behavior are. I report on a qualitative study aimed at answering these questions and discuss its results.

8.2 Prototype System

With a goal of generating a short story-like voice output about the current sales trends in a given category of product, I developed a prototype that takes as input a set of digested statistics about the sales and search trends, chooses what content to focus on, and outputs readable text describing the chosen trends. The system architecture can be seen in Fig. 8.1.

The *input trends* are retrieved from eBay's website and consist of data about the sales and searches of products. These data are digested and analyzed, previous to being used in the prototype, and can describe comparative statistics across time periods or products. For instance,

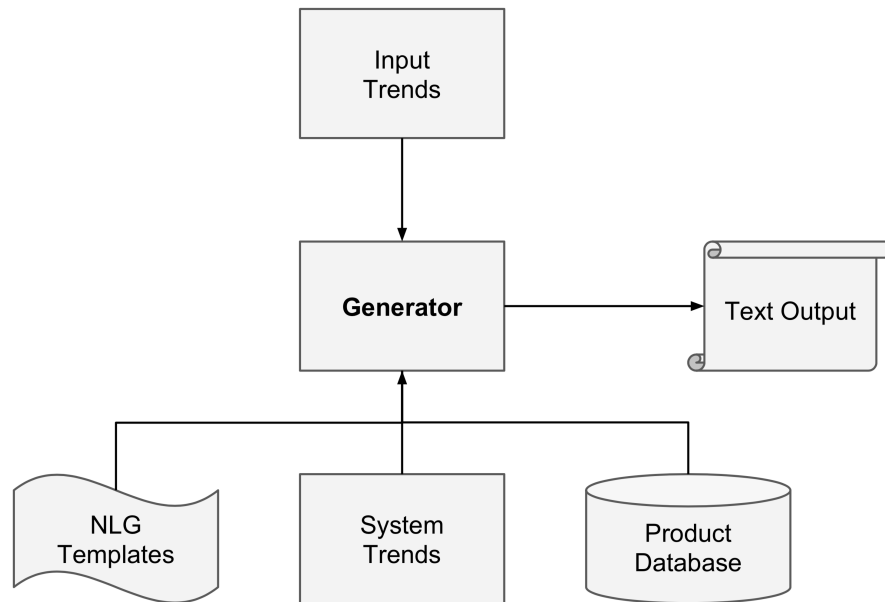


Figure 8.1: Prototype system's architecture.

input trends can point to a surge in the popularity of an item in terms of the number of times that users have searched for it in a given time period. Other examples include a list of the most popular items in terms of sales or a significant change in the price of a product (e.g., sales or discounts). *System trends* are a set of common input trends for which the prototype has pre-authored Natural Language Generation (NLG) templates.

Some of the corresponding system trends for the example input trends mentioned above are:

- ProductPopularitySurge
- MostPopularProductInCategory
- ProductDiscountTrend

Trends that are about a particular product are called *product trends*, and those that are more broadly about a category of product, are called *category trends*. These system trends each have qualifiers that determine the values for their associated variables, such as product IDs, time frame, or discount amounts. *Product database* has a list of products, their categories, brands, current price, and other relevant metadata, including a manually authored “design story” which highlights a background fact about a given product.

Given a product category and the input trends for that category, the generator will first look for known system trends among the input trends. Then, among the matches, the system looks for category trends and picks one at random. This category trend is used as the first part of the output. To form the second part, the system then focuses on product trends and finds the product that is most frequently the subject of trends (if multiple items are equally frequent, then it picks one at random). This product will become the *focus product* of the generation. Two product trends about the focus product are then chosen to follow the category trend. If no product has two or more trends associated with it, the two trends will be about different products in the chosen category. This process results in 3 total trends; one about the category and two about products.

Every system trend has an NLG template assigned to it. In the next step, these 3 trends are realized into text using the said templates and the trends’ associated metadata retrieved from the product database. Lastly, a design story about the focus product is incorporated after the first product trend (if one exists). Table 8.1 shows an example output of the system about Sneakers.

Table 8.1: Sample output of the system about Sneakers.

More sneakers dropped recently including Yeezy Boost 700 and Adidas Desert Rat Black (NewCategoryProductsTrend). The Adidas Desert Rat Black is the most trending Sneaker (MostPopularProductInCategory). Not just another basic black sneaker, the latest drop from Yeezy is a tonal mix of black mesh, black suede, and a black retro futuristic 1990s-inspired sole (Design Story). The popularity of Adidas Desert Rat Black has increased 30 percent since last month (ProductPopularitySurge).

8.3 Evaluation

To evaluate this approach and prototype, I then conducted a semi-structured qualitative user study. I recruited 9 subjects who were non-technical employees of eBay, including 5 females and 4 males; their average age was 29 (min: 21, max: 44).

8.3.1 Setup and Procedure

I used three product categories of sneakers, handbags, and drones to generate three sample outputs of the generator. In order to present these outputs to the participants, I used a platform called DialogFlow¹ which is capable of deploying voice experiences on a Google Home smart speaker. The experiences could then be invoked via an activation phrase, e.g. “demo number 3”. The DialogFlow experience included a short exchange for the user and started by asking them if they want to learn about what the agent can do. If answered positively by the user, the agent would give a short description of what it is capable of (“I can give you

¹<https://dialogflow.com/>

rundowns about what people are searching for, what they are buying, and what is popular in general”). Afterward, the agent offered the user to hear a sample, and upon acceptance by the user, the agent read the prototype output from one of the three product categories. A product category was randomly assigned to each of the 9 participants. Each category was used for 3 participants.

8.3.2 Results and Discussion

In this section, I will share the results of the study and discuss them in the context of my research motivations.

Participants who already do recreational browsing saw voice as an appropriate medium for receiving trend information.

Out of 9 participants, 4 of them (P1, P7, P8, and P9; all above 30 years of age) declared their shopping behavior to be largely need-based, while the rest of the participants (all below 30 years of age) said they more frequently browse shopping items for fun and look for trend information. The latter group mentioned web search and social media (e.g. Instagram) as their main sources of receiving trend information. P3 said: “*online, I go on Instagram a lot, I follow some influencers and look at where they buy their stuff;*” and P6, who browses items daily, said: “*I go on Google and look at the trends and price information [...] I’m also part of a Facebook group.*” The same participant later said about the voice demo: “*this can replace some of the browsings I do every day,*” and P4 said: “*[...] what are some fashion trends now? These are questions I like to know the answer to [...] I would be very interested to just use a Google*

home for this.”

Need-based shoppers see the value of the voice experience in doing research before buying.

Previously mentioned participants who are need-based shoppers saw the voice experience as a way to research a specific category of products that they have already decided to make a purchase from, by comparing prices and seeing what is popular. P7 said: *“I would use this for bigger things where you spend more money because I would research them more [...] I can get the information and just make the decision later whether I want to buy this or that.”* P8 and P9 also pointed to such comparative analyses before making a purchase.

This experience can support future purchases and potentially change user’s shopping behavior.

The behavior of need-based shoppers does not seem to be affected by the voice experience beyond helping with the research phase of buying, as mentioned in the previous finding. P1 said: *“it does not change the way I shop [...] if I need to buy something I buy it [...] I’m not a window shopper.”* Other participants, however, who were not purely need-based shoppers, believed having access to this experience to have the potential to change their shopping behavior. P2, for instance, said: *“it’d definitely get me on my phone to start searching for stuff.”* Moreover, multiple participants pointed out a need for finding the items that they would hear about in the voice experience, describing such functionality as “bookmarking” (P6) or adding items to a “wish list” (P5) as a part of the voice interaction. Giving the users this ability also enables a potential visual experience to follow the main voice experience at a later time, which

was noted as something desirable by 2 of the participants (P2 and P4).

Background information about products is desired by trend shoppers.

Participants who were not need-based shoppers appreciated the background information about products, describing it as interesting, while participants who were need-based shoppers did not care about such information. P3, talking about the design story of a handbag product which included information about a celebrity, said: *“I think the celebrity thing was cool when [the voice] was mentioning celebrities who like [the item].”* Meanwhile, P8 said of the same information: *“I mean I don’t really care personally [...] I’m not super brand driven.”*

Voice experience can benefit from including other hedonic motivations.

Even though I did not explicitly ask the participants about other hedonic motivations of shopping (see Sec. 8.1), they expressed their interest in receiving information about deals and discounts (value shopping), gift guides (role shopping), and shopping as a shared activity (social shopping). P6 said: *“I’d want more pricing [information], like trending low and trending high, or average maybe.”* P5 mentioned: *“when it transitioned to price drop it kind of got my attention.”* P8, talking about how they would use such voice experience, said: *“if I was trying to buy someone a gift [...] and it was an area where I wasn’t comfortable with, I could see myself [using] that information.”* And lastly, P9, whom I identified as a need-based shopper, saw this experience as an opportunity for having a shared activity at home: *“let’s say as a family we’re just sitting around, we don’t have to look at something all of us at the same screen [...], can just keep asking [via voice] and each one of us can take turns.”*

Voice can introduce shopping-related experiences to new contexts of usage.

Participants pictured themselves using this voice experience while doing activities that are not usually associated with shopping, such as “in the car” (P2), “in my morning news listening routine” (P5), “cooking” (P7), or “walking in the backyard” (P7).

It would be interesting to study how the voice experience can change and improve based on known user activity (e.g., smart speaker’s location), especially to avoid cognitive overload of the user as a result of multitasking.

Users strongly prefer to have an interactive experience.

Participants wanted the experience to be more interactive in order to control the flow of information and to guide and specify it easier. P4, for example, said: *“a more structured rundown [would be better:] ‘hey these are the most popular things, let me know if you want to hear more about a specific shoe’ [...] then I can say ‘tell me more about the Yeezys’.”* P8, who is a more need-based shopper, wanted even more control and said: *“I would want to be able to ask the specific question that I want the answer to rather than [...] getting a list of information, a more targeted question”*.

As a need-based shopper, P8’s desire for detailed question answering is compatible with a tendency to use the experience for researching before buying. Hence, it would be interesting to predict what styles of interactivity might improve a given user’s experience on that basis.

The topics covered in the experience can be expanded.

Participants mentioned other topics than trends that they would want to hear about in such experience, including “news about shops and brands” (P4) and “keywords of product reviews” (P6). Interestingly, P2 mentioned that they would use this experience for receiving content recommendations as well, such as getting information about popular “documentaries”.

8.4 Chapter Takeaways

Online shopping and usage of voice-enabled devices are both widespread, but it is not clear if buying through voice is becoming more popular. Hedonic motivations of shopping may introduce an opportunity for shopping-related voice experiences that are not directly advertisements or about the act of purchasing but are about the interest in shopping; analogous to exploration in a mall for the joy of it alone. This introduces an opportunity for story generation. In this work, through an internship at eBay, the popular online shopping platform, I created a voice prototype that focuses on the hedonic aspects of shopping by providing shopping trend information to the users in a short story-like experience. This story snippet focuses on topic interests from experiential interests (see chapter 3). I reported on an evaluation, semi-structured qualitative interviews, with a goal of better understanding this new type of experience and deriving design recommendations.

Chapter 9

Conclusions

The interest in story generation is increasing, the set of approaches taken to generate stories is expanding, and these approaches are themselves becoming more successful with the advances made in various methods. With the said increase, expansion and improvement, and along with other technological advances such as ubiquitous computing, interactive agents and sociable robots [90, 105], comes emerging new use cases for story generation. In order to support such use cases, I believe that the generation of stories should be made aware of the cognitive processes involved in the perception of them.

In this dissertation, I have presented work on making the computational story generation, and in particular, open story generation, aware of the cognitive processes involved in the perception of stories, with a goal of curating more interest. In the first part of this dissertation, titled “The Why”, I outlined the reasoning behind the need for a new focus on the perception of stories in the exercise of generating them computationally.

In chapter 2, I focused on the role of stories in human life and the root of their prevalence in all aspects of the human experience. In the same chapter, I then focused on an entanglement that is acknowledged and studied in psychology, cognitive science, and narratology. In this process, the human storyteller, whether she is aware of it or not, takes a listener's perception into consideration as she tells a story, involving a process of theory of mind [76, 125]. It is apparent, from this literature, that a capable story generator should do the same.

To this end, in chapter 3, I focused on an understanding of the cognitive processes involved in the perception of stories and attempted to answer the simple question: "what makes stories be perceived as interesting?". Drawing from psychology and cognitive science literature, I consolidated the knowledge in this space with a focus on the viability of their use in computational story generation. I then expanded these theories and provided a partial evaluation of it inspired by the classic Fritz Heider and Mary-Ann Simmel [75] experiments on apparent behavior. This provided answers for my RQ3 discussed in chapter 1, resulting in a usable theory.

Importantly, a multitude of different approaches are taken to narrative modeling and story generation, and hence, this direction of work is heavily affected by the level of reliance that such approaches have on assuming the semantics of a particular domain. On one end, some approaches focus on heavily modeling one specific domain and then enabling a story generator to tell stories about that domain; such as it is often the case in games. However, the future use cases of story generation, such as interactions with an agent that tells situated stories, do not lend themselves well to these approaches. Hence, I focused on the other end of this spectrum, on *open story generation*, in which there exists an emphasis on generating stories without such a priori domain semantics. Crucially, the assumption of semantics also affects the approaches

that can be taken to evaluate the generated stories. In the “heavily-modeled” end, it is relatively much easier to use the domain semantics to evaluate the generated stories: a story involving a checkmate is likely to be more interesting than a story about random other moves in chess. On the other end, the evaluation metrics currently used in the domain of open story generation are neither aware of the qualities of the perception nor are they metrics created for the evaluation of *stories*; instead, they focus on language properties such as text coherence that are not necessarily appropriate or complete qualifiers for the goodness of stories. I offered a discussion of these nuances and a report on the relevant related works in chapter 4. This provided answers for my RQ1 discussed in chapter 1, as I learned about the requirements and consequences of various approaches to generating stories.

In the second part of this dissertation, titled “The How”, I offered two approaches that satisfy the needs discussed in the first part.

In chapter 5, I focused on the generation of stories without a priori models, and particularly, with an approach that lends itself well to a future use case in which an interactive agent creates situated stories from observed or experienced events. I then provided a small user study to evaluate this approach. While other approaches can be taken to open story generation, as discussed before, what is crucial is the evaluation of their generated stories with a focus on perception. However, an understanding of these processes and even a computationally viable theory is not enough. Specific quantitative measures are needed to estimate the perception of stories in order to easily incorporate them in a story generation process. This provided answers for my RQ2 discussed in chapter 1.

Hence, in chapter 6, informed by the theories discussed earlier in the dissertation, I

reported on the creation of a proxy measure using word embeddings and laid out a two-phased evaluation that shows the proxy measure to be correlated with human judgment. This provided answers for my RQ4 discussed in chapter 1.

Lastly, in the third part of the dissertation, titled “Stories in the Wild”, I reported on the results of my internships at companies Spotify and eBay which led to the creation of two prototypes that use story generation techniques with influences from the theories of interestingness discussed in chapter 3. In both cases, I reported on an evaluation in the form of semi-structured qualitative interviews. This exploration provided additional answers and insights for RQ4.

9.1 Discussion

The presented perception-focused story quality proxy measure can connect well with the open story generation approaches; including the one presented in chapter 5, through ranking the suggested stories in that method based on the proxy measure’s value. However, other approaches to open story generation, such as those involving neural networks, can benefit from this proxy measure as well. In particular, sampling from trained models, either conditioned on an input text or not, can sometimes generate surprisingly coherent snippets of text. OpenAI’s GPT2 model [138], for instance, is a Language Model that has produced promising results. With this observation, and as simpler downstream tasks in the field of NLP pose fewer challenges than before, there has been a surge in using such approaches to generate stories.

However, there are paradoxes in this effort. Sampling on trained neural networks, either conditioned on an input or not, is a distinctly different process than storytelling in that it

is not controlled or optimized to be a “story”, much less an interesting one. Hence, researchers and practitioners often generate many samples and report on ones that, *incidentally*, happen to resemble coherence or story interestingness. It is surely plausible that with rapid advances in this area we will see much better generated text *more often*, however, it is not clear that this approach by itself will guide us to reliably generating good stories. Moreover, one would supposedly intend to use these generated stories in a particular interaction or situation, and thus, a control in the generation of stories in this approach is key. New research on using reinforcement learning to achieve a controllable neural story generation [169] is a promising approach, and introduces an appropriate point for building a bridge to proxy measures of perception of stories.

9.2 Recommendation for Future Work

I believe that my work on understanding the cognitive processes involved in the perception of stories and on the creation of a proxy measure for cognitive interest in stories represents an important and useful bridge in the domain of story generation. But crucially, this is only a starting point for a new direction of research.

Creating *cognitive models* of the perception of stories is an intriguing direction that can incorporate more aspects of perception, cognition, and situation. For instance, such cognitive models can be combined with user models, memory, and affect models to include experiential interests reported in chapter 3. Hence, creating explicit cognitive models of comprehension of narrative and stories is an interesting avenue to pursue, one that can potentially affect the

machine learning and neural generation of stories as well, by providing direction and a more applicable pallet of evaluation.

As experiential interests are more subjective, modeling them opens the door to the possibility of creating personalized story interestingness models. While many aspects of the perception of stories (such as unexpectedness or predictive inference) are valid across humans, as mentioned in chapter 3, story interests are likely to interact in the human mind, and hence, a more comprehensive cognitive model can also address personalization more effectively.

As computing becomes more ubiquitous, intelligent, and interactive, story generation could become an integral part of a future where agents use natural interfaces such as language to interact with humans. Some such agents, as evident in the field of Human-Robot Interaction, will benefit from exhibiting social cues and even creating a social rapport and relationship with the user [90, 105, 158]. Due to the prevalence of stories in our culture, socialization norms, and cognition, I believe storytelling will have a great role in the future of sociable interactive agents. To this end, it is worth investigating the effects of the properties of an agent, including its embodiment, level of anthropomorphism, its relationship with the human counterparts, and so on, on the perception of the stories it communicates. In turn, this knowledge can be used to make the generated stories better and derive recommendations for various agents and contexts.

Lastly, I believe that my experience in using story generation in vastly different use cases than games, as it is reported in part III of this dissertation, and my observation of perceived interest in such generated stories shines a light on the possibility of expanding the use cases of story generation much further.

Bibliography

- [1] Dominic Abrams and Michael A Hogg. *Social identifications: A social psychology of intergroup relations and group processes*. Routledge, 2006.
- [2] Henny Admoni and Brian Scassellati. Social eye gaze in human-robot interaction: a review. *Journal of Human-Robot Interaction*, 6(1):25–63, 2017.
- [3] Nicole Alea and Susan Bluck. Why are you telling me that? a conceptual model of the social function of autobiographical memory. *Memory*, 11(2):165–178, 2003.
- [4] Nicholas D Allen, John R Templon, Patrick Summerhays McNally, Larry Birnbaum, and Kristian J Hammond. Statsmonkey: A data-driven sports narrative writer. In *AAAI Fall Symposium: Computational Models of Narrative*, 2010.
- [5] Music Ally. Everybody’s talkin’: Smart speakers & their impact on music consumption. <http://musically.com/wp-content/uploads/2018/03/SmartSpeakersFinal.pdf>, 2018. Accessed: 2018-07-20.
- [6] Anupriya Ankolekar, Thomas Sandholm, and Louis Lei Yu. Evaluating mobile music experiences: Radio on-the-go. In *International Conference on Mobile Computing, Applications, and Services*, pages 56–73. Springer, 2018.
- [7] AppleInsider. Photos memories will generate slideshow movies automatically. <https://appleinsider.com/articles/16/06/16/inside-ios-10-photos-memories-will-generate-slideshow-movies-automatically>, 2016. Accessed: 2018-07-25.
- [8] Rudolf Arnheim. *Art and visual perception: A psychology of the creative eye*. Univ of California Press, 1956.
- [9] Mark J Arnold and Kristy E Reynolds. Hedonic shopping motivations. *Journal of retailing*, 79(2):77–95, 2003.
- [10] Byung-Chull Bae and R Michael Young. A use of flashback and foreshadowing for surprise arousal in narrative using a plan-based approach. In *Joint International Conference on Interactive Digital Storytelling*, pages 156–167. Springer, 2008.

- [11] David Bainbridge, Sally Jo Cunningham, and J. Stephen Downie. How people describe their music information needs: A grounded theory analysis of music queries. In *ISMIR*, pages 221–222, 2003.
- [12] Mieke Bal and Christine Van Boheemen. *Narratology: Introduction to the theory of narrative*. University of Toronto Press, 2009.
- [13] Roland Barthes and Lionel Duisit. An introduction to the structural analysis of narrative. *New literary history*, 6(2):237–272, 1975.
- [14] Cristina Battaglino and Timothy Bickmore. Increasing the engagement of conversational agents through co-constructed storytelling. In *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference*, 2015.
- [15] Roy F Baumeister and Leonard S Newman. How stories make sense of personal experiences: Motives that shape autobiographical narratives. *Personality and Social Psychology Bulletin*, 20(6):676–690, 1994.
- [16] BBC. Bbc launches first interactive voice drama. <http://www.bbc.co.uk/mediacentre/latestnews/2017/inspection-chamber>, 2017. Accessed: 2018-07-01.
- [17] Morteza Behrooz, Charles Rich, and Candace Sidner. On the sociability of a game-playing agent: A software framework and empirical study. In *Intelligent Virtual Agents*, pages 40–53. Springer, 2014.
- [18] Jody Berland. Radio space and industrial time: music formats, local narratives and technological mediation 1. *Popular Music*, 9(2):179–192, 1990.
- [19] Daniel E Berlyne. Conflict, arousal, and curiosity. 1960.
- [20] Daniel E Berlyne. Novelty, complexity, and hedonic value. *Attention, Perception, & Psychophysics*, 8(5):279–286, 1970.
- [21] DE Berlyne. Uncertainty and epistemic curiosity. *British Journal of Psychology*, 53(1):27–34, 1962.
- [22] Hugh Beyer and Karen Holtzblatt. *Contextual design: defining customer-centered systems*. Elsevier, 1997.
- [23] Timothy Bickmore and Julie Cassell. Small talk and conversational storytelling in embodied conversational interface agents. In *AAAI fall symposium on narrative intelligence*, pages 87–92, 1999.
- [24] Billboard. From mood playlists to metadata: How smart speakers are the next frontier – and challenge – for the music business. <https://www.billboard.com/articles/business/8263197/smart-speaker-challenges-music-business>, 2018. Accessed: 2018-07-20.

- [25] Peter H Bloch, Nancy M Ridgway, and Daniel L Sherrell. Extending the concept of shopping: An investigation of browsing activity. *Journal of the Academy of Marketing Science*, 17(1):13–21, 1989.
- [26] Peter H Bloch, Daniel L Sherrell, and Nancy M Ridgway. Consumer search: An extended framework. *Journal of consumer research*, 13(1):119–126, 1986.
- [27] Stefano Bocconi, Frank Nack, and Lynda Hardman. Automatic generation of matter-of-opinion video documentaries. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(2):139–150, 2008.
- [28] Nadjat Bouayad-Agha, Gerard Casamayor, and Leo Wanner. Content selection from an ontology-based knowledge base for the generation of football summaries. In *Proceedings of the 13th European Workshop on Natural Language Generation*, pages 72–81. Association for Computational Linguistics, 2011.
- [29] Gordon H Bower, John B Black, and Terrence J Turner. Scripts in memory for text. *Cognitive psychology*, 11(2):177–220, 1979.
- [30] William F Brewer and Edward H Lichtenstein. Event schemas, story schemas, and story grammars. *Center for the Study of Reading Technical Report; no. 197*, 1980.
- [31] Alan S Brown. A review of the déjà vu experience. *Psyc. bulletin*, 129(3):394, 2003.
- [32] Steven Brown, Michael J Martinez, and Lawrence M Parsons. Passive music listening spontaneously engages limbic and paralimbic systems. *Neuroreport*, 15(13):2033–2037, 2004.
- [33] Jerome Bruner. The narrative construction of reality. *Critical inquiry*, 18(1):1–21, 1991.
- [34] Jerome S Bruner. *Acts of meaning*, volume 3. Harvard University Press, 1990.
- [35] Eric Buckthal and Foaad Khosmood. (re)telling chess stories as game content. In *9th International Conference on the Foundations of Digital Games*, 2014.
- [36] Vivien Burr. *Social constructionism*. Routledge, 2015.
- [37] Nicolas Campion, Daniel Martins, and Alice Wilhelm. Contradictions and predictions: Two sources of uncertainty that raise the cognitive interest of readers. *Discourse Processes*, 46(4):341–368, 2009.
- [38] Rogelio E Cardona-Rivera and Boyang Li. Plotshot: Generating discourse-constrained stories around photos. In *Proceedings of the 12th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 2016.
- [39] Seymour Benjamin Chatman. *Story and discourse: Narrative structure in fiction and film*. Cornell University Press, 1980.

- [40] Ana Paula Chaves and Marco Aurelio Gerosa. Single or multiple conversational agents?: An interactional coherence comparison. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 191. ACM, 2018.
- [41] David L Chen and William B Dolan. Collecting highly parallel data for paraphrase evaluation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 190–200. Association for Computational Linguistics, 2011.
- [42] Terry L Childers, Christopher L Carr, Joann Peck, and Stephen Carson. Hedonic and utilitarian motivations for online retail shopping behavior. *Journal of retailing*, 77(4):511–535, 2001.
- [43] Clark. Voice shopping is the next big thing in retail. <https://clark.com/shopping-retail/voice-shopping-amazon-alexa-echo-google-home/>, 2018. Accessed: 2018-08-27.
- [44] Elizabeth Clark, Yangfeng Ji, and Noah A Smith. Neural text generation in stories using entity representations as context. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2250–2260, 2018.
- [45] Neil Cohn, Tom Foulsham, Tim J Smith, and Jeffrey M Zacks. Static and dynamic visual narratives, by brain and by eye. The Cognitive Science Society, 2017.
- [46] William Coon, Charles Rich, and Candace L Sidner. Activity planning for long-term relationships. In *Intelligent Virtual Agents: 13th International Conference, IVA 2013, Edinburgh, UK, August 29-31, 2013, Proceedings*, volume 8108, page 425. Springer, 2013.
- [47] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [48] Frank M Diaz. Mindfulness, attention, and flow during music listening: An empirical investigation. *Psychology of Music*, 41(1):42–58, 2013.
- [49] Belén Díaz-Agudo, Pablo Gervás, and Federico Peinado. A case based reasoning approach to story plot generation. In *European Conference on Case-Based Reasoning*, pages 142–156. Springer, 2004.
- [50] Giel Dik and Henk Aarts. Behavioral cues to others’ motivation and goal pursuits: The perception of effort facilitates goal inference and contagion. *Journal of Experimental Social Psychology*, 43(5):727–737, 2007.
- [51] Robin IM Dunbar. Gossip in evolutionary perspective. *Review of general psychology*, 8(2):100–110, 2004.

- [52] David K Elson. *Modeling narrative discourse*. 2012.
- [53] eMarketer. emarketer releases new smart speaker figures. <https://www.emarketer.com/content/emarketer-releases-new-smart-speaker-figures>, 2018. Accessed: 2018-08-27.
- [54] Susan Engel. *The stories children tell: Making sense of the narratives of childhood*. Macmillan, 1995.
- [55] Angela Fan, Mike Lewis, and Yann Dauphin. Hierarchical neural story generation. *arXiv preprint arXiv:1805.04833*, 2018.
- [56] Rachelyn Farrell and Stephen G Ware. Fast and diverse narrative planning through novelty pruning. In *Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2016.
- [57] Leon Festinger. A theory of social comparison processes. *Human relations*, 7(2):117–140, 1954.
- [58] Peter W Foltz, Walter Kintsch, and Thomas K Landauer. The measurement of textual coherence with latent semantic analysis. *Discourse processes*, 25(2-3):285–307, 1998.
- [59] Tom Foulsham, Dean Wybrow, and Neil Cohn. Reading without words: Eye movements in the comprehension of comic strips. *Applied Cognitive Psychology*, 30(4):566–579, 2016.
- [60] Philippe Fournier-Viger, Usef Faghihi, Roger Nkambou, and Engelbert Mephu Nguifo. Cmrules: Mining sequential rules common to several sequences. *Knowledge-Based Systems*, 25(1):63–76, 2012.
- [61] Peter Freebody and Richard C Anderson. Serial position and rated importance in the recall of text. *Discourse processes*, 9(1):31–36, 1986.
- [62] Robert W Frick. Interestingness. *British Journal of Psychology*, 83(1):113–128, 1992.
- [63] Albert Gatt and Ehud Reiter. Simplenlg: A realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 90–93. Association for Computational Linguistics, 2009.
- [64] Richard J Gerrig. *Experiencing narrative worlds: On the psychological activities of reading*. Yale University Press, 1993.
- [65] Pablo Gervás, Belén Díaz-Agudo, Federico Peinado, and Raquel Hervás. Story plot generation based on cbr. *Knowledge-Based Systems*, 18(4):235–242, 2005.
- [66] Anna Gladkova, Aleksandr Drozd, and Satoshi Matsuoka. Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn’t. In *Proceedings of the NAACL Student Research Workshop*, pages 8–15, 2016.

- [67] Yoav Goldberg and Omer Levy. word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*, 2014.
- [68] Paulo Gomes, Ana Paiva, Carlos Martinho, and Arnav Jhala. Metrics for character believability in interactive narrative. In *International Conference on Interactive Digital Storytelling*, pages 223–228. Springer, 2013.
- [69] Google. Travel through hawkins with new stranger things game for google home. <https://blog.google/products/home/travel-through-hawkins-new-stranger-things-game-google-home/>, 2017. Accessed: 2018-07-01.
- [70] Jonathan Gottschall. *The storytelling animal: How stories make us human*. Houghton Mifflin Harcourt, 2012.
- [71] Arthur C Graesser, Danielle S McNamara, Max M Louwerse, and Zhiqiang Cai. Coh-metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers*, 36(2):193–202, 2004.
- [72] Melanie C Green. Transportation into narrative worlds: The role of prior knowledge and perceived realism. *Discourse processes*, 38(2):247–266, 2004.
- [73] Mitch Griffin, Barry J Babin, and Doan Modianos. Shopping values of russian consumers: the impact of habituation in a developing economy. *Journal of Retailing*, 76(1):33–52, 2000.
- [74] Andrew F Hayes and Klaus Krippendorff. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1):77–89, 2007.
- [75] Fritz Heider and Marianne Simmel. An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2):243–259, 1944.
- [76] David Herman. *Storytelling and the Sciences of Mind*. MIT press, 2013.
- [77] Marcia Herndon and Norma McLeod. *Music as culture*. Norwood, 1981.
- [78] Suzanne Hidi and William Baird. Interestingness—a neglected variable in discourse processing. *Cognitive Science*, 10(2):179–194, 1986.
- [79] Christine Hosey, Lara Vujovic, Brian St. Thomas, Jean Garcia-Gathright, and Jennifer Thom. Just give me what I want: How people use and evaluate music search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2019.
- [80] Jeremy Hsu. The secrets of storytelling: Why we love a good yarn. *Scientific American Mind*, 19(4):46–51, 2008.

- [81] Business Insider. Smart speakers are becoming so popular, more people will use them than wearable tech products this year. <https://www.businessinsider.com/more-us-adults-will-use-smart-speakers-than-wearables-in-2018-2018-5>, 2018. Accessed: 2018-08-27.
- [82] Asghar Iran-Nejad. Cognitive and affective causes of interest and liking. *Journal of Educational Psychology*, 79(2):120, 1987.
- [83] Fred Jelinek, Robert L Mercer, Lalit R Bahl, and James K Baker. Perplexity—a measure of the difficulty of speech recognition tasks. *The Journal of the Acoustical Society of America*, 62(S1):S63–S63, 1977.
- [84] Hua Jin, Ho-Ling Liu, Lei Mo, Shin-Yi Fang, John X Zhang, and Chong-De Lin. Involvement of the left inferior frontal gyrus in predictive inference making. *International Journal of Psychophysiology*, 71(2):142–148, 2009.
- [85] Armand Joulin, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Hérve Jégou, and Tomas Mikolov. Fasttext.zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651*, 2016.
- [86] Michael Joyce. Afternoon, a story.[cd-rom]. *Watertown, MA: Eastgate Systems*, pages 579–97, 1987.
- [87] Keith Kahn-Harris. *Extreme metal: Music and culture on the edge*. Berg, 2006.
- [88] Harold H Kelley. The processes of causal attribution. *American psychologist*, 28(2):107, 1973.
- [89] Ahmed Khalifa, Gabriella AB Barros, and Julian Togelius. Deeptingle. *arXiv preprint arXiv:1705.03557*, 2017.
- [90] Elizabeth S Kim, Lauren D Berkovits, Emily P Bernier, Dan Leyzberg, Frederick Shic, Rhea Paul, and Brian Scassellati. Social robots as embedded reinforcers of social behavior in children with autism. *Journal of autism and developmental disorders*, 43(5):1038–1049, 2013.
- [91] Hyejeong Kim, Ann Marie Fiore, Linda S Niehm, and Miyoung Jeong. Psychographic characteristics affecting behavioral intentions towards pop-up retail. *International Journal of Retail & Distribution Management*, 38(2):133–154, 2010.
- [92] Sung-il Kim. Causal bridging inference: A cause of story interestingness. *British Journal of Psychology*, 90(1):57–71, 1999.
- [93] Walter Kintsch. Learning from text, levels of comprehension, or: Why anyone would read a story anyway. *Poetics*, 9(1-3):87–98, 1980.
- [94] C Klimas. Twine: an open-source tool for telling interactive, nonlinear stories.

- [95] Stefan Kopp, Lars Gesellensetter, Nicole C Krämer, and Ipke Wachsmuth. A conversational agent as museum guide—design and evaluation of a real-world application. In *Intelligent virtual agents*, pages 329–343. Springer, 2005.
- [96] William Labov and Joshua Waletzky. Narrative analysis: Oral versions of personal experience. 1997.
- [97] François Lareau, Mark Dras, and Robert Dale. Detecting interesting event sequences for sports reporting. In *Proceedings of the 13th European Workshop on Natural Language Generation*, pages 200–205. Association for Computational Linguistics, 2011.
- [98] Richard S Lazarus. *Emotion and adaptation*. Oxford University Press on Demand, 1991.
- [99] Michael Lebowitz. Planning stories. In *Proceedings of the 9th annual conference of the cognitive science society*, pages 234–242, 1987.
- [100] Jin Ha Lee, Hyerim Cho, and Yea-Seul Kim. Users’ music information needs and behaviors: Design implications for music information retrieval systems. *Journal of the association for information science and technology*, 67(6):1301–1330, 2016.
- [101] Jin Ha Lee, Yea-Seul Kim, and Chris Hubbles. A look at the cloud from both sides now: An analysis of cloud music service usage. In *ISMIR*, pages 299–305, 2016.
- [102] Jin Ha Lee and Rachel Price. Understanding users of commercial music services through personas: Design implications. In *ISMIR*, pages 476–482, 2015.
- [103] Kun Chang Lee and Namho Chung. Empirical analysis of consumer reaction to the virtual reality shopping mall. *Computers in Human Behavior*, 24(1):88–104, 2008.
- [104] Moritz Lehne, Philipp Engel, Martin Rohrmeier, Winfried Menninghaus, Arthur M Jacobs, and Stefan Koelsch. Reading a suspenseful literary text activates brain areas related to social cognition and predictive inference. *PLoS One*, 10(5):e0124550, 2015.
- [105] Iolanda Leite, Carlos Martinho, and Ana Paiva. Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, 5(2):291–308, 2013.
- [106] Ang Li, Jennifer Thom, Praveen Chandar, Christine Hosey, Brian St. Thomas, and Jean Garcia-Gathright. Search mindsets: Understanding focused and non-focused information seeking in music search. In *Proceedings of the 30th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 2019.
- [107] Boyang Li, Stephen Lee-Urban, George Johnston, and Mark Riedl. Story generation with crowdsourced plot graphs. In *AAAI*, 2013.
- [108] Adam J Lonsdale and Adrian C North. Why do we listen to music? a uses and gratifications analysis. *British Journal of Psychology*, 102(1):108–134, 2011.

- [109] Stephanie M Lukin, Lena I Reed, and Marilyn A Walker. Generating sentence planning variations for story telling. *arXiv preprint arXiv:1708.08580*, 2017.
- [110] George Mandler. The structure of value: Accounting for taste. *Center for Human Information Processing Report*, 101, 1982.
- [111] Inderjeet Mani. Computational narratology. *Handbook of narratology*, pages 84–92, 2014.
- [112] Raymond A Mar and Keith Oatley. The function of fiction is the abstraction and simulation of social experience. *Perspectives on psychological science*, 3(3):173–192, 2008.
- [113] Lara J Martin, Prithviraj Ammanabrolu, Xinyu Wang, William Hancock, Shruti Singh, Brent Harrison, and Mark O Riedl. Event representations for automated story generation with deep neural nets. *arXiv preprint arXiv:1706.01331*, 2017.
- [114] Michael Mateas and Andrew Stern. Façade: An experiment in building a fully-realized interactive drama. In *Game developers conference*, volume 2, pages 4–8, 2003.
- [115] Philip J Mazzocco, Melanie C Green, Jo A Sasota, and Norman W Jones. This story is not for everyone: Transportability and narrative persuasion. *Social Psychological and Personality Science*, 1(4):361–368, 2010.
- [116] Josh McCoy, Mike Treanor, Ben Samuel, Michael Mateas, and Noah Wardrip-Fruin. Prom week: social physics as gameplay. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, pages 319–321. ACM, 2011.
- [117] William J McGuire. Psychological motives and communication gratification. *The uses of mass communications: Current perspectives on gratifications research*, 3:167–196, 1974.
- [118] Kate C McLean, Monisha Pasupathi, and Jennifer L Pals. Selves creating stories creating selves: A process model of self-development. *Personality and Social Psychology Review*, 11(3):262–278, 2007.
- [119] James R Meehan. Tale-spin, an interactive program that writes stories. In *IJCAI*, volume 77, pages 91–98, 1977.
- [120] George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
- [121] Esa Nettamo, Mikko Nirhamo, and Jonna Häkkinen. A cross-cultural study of mobile music: retrieval, management and consumption. In *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, pages 87–94. ACM, 2006.

- [122] Market Communication News. One in five use alexa to boil an egg, according to code computerlove report. <http://marcommnews.com/one-in-five-use-alexa-to-boil-an-egg-according-to-code-computerlove-report/>, 2018. Accessed: 2018-08-27.
- [123] Adrian C North and David J Hargreaves. Situational influences on reported musical preference. *Psychomusicology: A Journal of Research in Music Cognition*, 15(1-2):30, 1996.
- [124] Edison Research NPR. The smart audio report. <https://www.nationalpublicmedia.com/smart-audio-report/>, 2017. Accessed: 2018-07-20.
- [125] Keith Oatley. Fiction: Simulation of social worlds. *Trends in cognitive sciences*, 20(8):618–628, 2016.
- [126] Daniela K O’Neill and Rebecca M Shultis. The emergence of the ability to track a character’s mental perspective in narrative. *Developmental Psychology*, 43(4):1032, 2007.
- [127] Santiago Ontañón and Jichen Zhu. Story and text generation through computational analogy in the riu system. In *Sixth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2010.
- [128] Matteo Pagliardini, Prakhar Gupta, and Martin Jaggi. Unsupervised learning of sentence embeddings using compositional n-gram features. *arXiv preprint arXiv:1703.02507*, 2017.
- [129] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics, 2002.
- [130] Andrew G Parsons. Non-functional motives for online shoppers: why we click. *Journal of Consumer Marketing*, 19(5):380–392, 2002.
- [131] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.
- [132] Rafael Pérez Y Pérez and Mike Sharples. Mexica: A computer model of a cognitive account of creative writing. *Journal of Experimental & Theoretical Artificial Intelligence*, 13(2):119–139, 2001.
- [133] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*, 2018.

- [134] Livia Polanyi. *Telling the American story: A structural and cultural analysis of conversational storytelling*. Ablex Publishing Corporation, 1985.
- [135] Julie Porteous, Marc Cavazza, and Fred Charles. Applying planning to interactive storytelling: Narrative control using state constraints. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 1(2):10, 2010.
- [136] Mary Louise Pratt. *Toward a speech act theory of literary discourse*. Indiana Univ Pr, 1977.
- [137] Christopher Purdy, Xinyu Wang, Larry He, and Mark Riedl. Predicting generated story quality with quantitative measures. 2018.
- [138] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.
- [139] David N Rapp and Richard J Gerrig. Predilections for narrative outcomes: The impact of story contexts and reader preferences. *Journal of Memory and Language*, 2006.
- [140] Mark O Riedl and R Michael Young. An objective character believability evaluation procedure for multi-agent story generation systems. In *Intelligent Virtual Agents*, pages 278–291, 2005.
- [141] Mark O Riedl and R Michael Young. From linear story generation to branching story graphs. *Computer Graphics and Applications, IEEE*, 26(3):23–31, 2006.
- [142] Mark O Riedl and Robert Michael Young. Narrative planning: Balancing plot and character. *Journal of Artificial Intelligence Research*, 39(1):217–268, 2010.
- [143] Mark Owen Riedl and Vadim Bulitko. Interactive narrative: An intelligent systems approach. *AI Magazine*, 34(1):67, 2012.
- [144] Melissa Roemmele. *Neural Networks for Narrative Continuation*. PhD thesis, University of Southern California, 2018.
- [145] Melissa Roemmele, Sosuke Kobayashi, Naoya Inoue, and Andrew Gordon. An rnn-based binary classifier for the story cloze test. In *Proceedings of the 2nd Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics*, pages 74–80, 2017.
- [146] Tricia Rose. *Black noise: Rap music and black culture in contemporary America*, volume 6. Wesleyan University Press Middletown, CT, 1994.
- [147] Jonathan P Rowe, Lucy R Shores, Bradford W Mott, and James C Lester. Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of Artificial Intelligence in Education*, 21(1-2):115–133, 2011.
- [148] James Ryan. *Curating Simulated Storyworlds*. PhD thesis, UC Santa Cruz, 2018.

- [149] Theodore R Sarbin. *Narrative psychology: The storied nature of human conduct*. Praeger Publishers/Greenwood Publishing Group, 1986.
- [150] Daniele Scarpi. Work and fun on the internet: the effects of utilitarianism and hedonism online. *Journal of interactive marketing*, 26(1):53–67, 2012.
- [151] Daniele Scarpi, Gabriele Pizzi, and Marco Visentin. Shopping for fun or shopping to buy: Is it different online and offline? *Journal of Retailing and Consumer Services*, 21(3):258–267, 2014.
- [152] Roger C Schank. Interestingness: controlling inferences. *Artificial intelligence*, 12(3):273–297, 1979.
- [153] Roger C Schank and Robert P Abelson. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Psychology Press, 2013.
- [154] Karin Kipper Schuler. Verbnet: A broad-coverage, comprehensive verb lexicon. 2005.
- [155] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I Hong. Hey alexa, what’s up?: A mixed-methods studies of in-home conversational agent usage. In *Proceedings of the Designing Interactive Systems Conference*, pages 857–868. ACM, 2018.
- [156] John R. Searle. A taxonomy of illocutionary acts. In Keith Gunderson, editor, *Language, Mind and Knowledge*, pages 344–369. University of Minnesota Press, 1975.
- [157] Murray Shanahan. The event calculus explained. In *Artificial intelligence today*, pages 409–430. Springer, 1999.
- [158] Candace Sidner, Timothy Bickmore, Charles Rich, Barbara Barry, Lazlo Ring, Morteza Behrooz, and Mohammad Shayganfar. An always-on companion for isolated older adults. In *14th Annual SIGdial meeting on discourse and dialogue*, 2013.
- [159] Candace L Sidner. Engagement, emotions, and relationships: On building intelligent agents. *Emotions, Technology, Design, and Learning*, page 273, 2015.
- [160] Paul J Silvia. What is interesting? exploring the appraisal structure of interest. *Emotion*, 5(1):89, 2005.
- [161] Paul J Silvia. *Exploring the psychology of interest*. Oxford University Press, 2006.
- [162] Paul J Silvia. Interest the curious emotion. *Current Directions in Psychological Science*, 17(1):57–60, 2008.
- [163] Statista. Online-shopping and e-commerce worldwide: Statistics & facts. <https://www.statista.com/topics/871/online-shopping/>, 2018. Accessed: 2018-08-16.
- [164] Statista. Online shopping behavior in the united states - statistics & facts. <https://www.statista.com/topics/2477/online-shopping-behavior/>, 2018. Accessed: 2018-08-16.

- [165] Statista. Retail e-commerce sales worldwide. <https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/>, 2018. Accessed: 2018-08-16.
- [166] Elliot Jay Stocks. Music consumption in the era of smart speakers. <https://medium.com/@elliottjaystocks/music-consumption-in-the-era-of-smart-speakers-b88d04a18746>, 2017. Accessed: 2018-07-20.
- [167] Reid Swanson and Andrew S Gordon. Say anything: A massively collaborative open domain story writing companion. In *Interactive Storytelling*, pages 32–40. Springer, 2008.
- [168] Reid Swanson and Andrew S Gordon. Say anything: Using textual case-based reasoning to enable open-domain interactive storytelling. *ACM Transactions on Interactive Intelligent Systems (TiIS)*, 2(3):16, 2012.
- [169] Pradyumna Tambwekar, Murtaza Dhuliawala, Animesh Mehta, Lara J Martin, Brent Harrison, and Mark O Riedl. Controllable neural story generation via reinforcement learning. *arXiv preprint arXiv:1809.10736*, 2018.
- [170] Edward M Tauber. Why do people shop? *The Journal of Marketing*, pages 46–49, 1972.
- [171] TechCrunch. Instagram adds shopping tags directly into stories. <https://techcrunch.com/2018/06/12/instagram-adds-shopping-tags-directly-into-stories/>, 2018. Accessed: 2018-08-27.
- [172] TechCrunch. Amazon to control 70 percent of the voice-controlled speaker market this year. <https://techcrunch.com/2017/05/08/amazon-to-control-70-percent-of-the-voice-controlled-speaker-market-this-year/>, 2017. Accessed: 2018-08-27.
- [173] TechCrunch. Google photos can now automatically create themed movies on demand. <https://techcrunch.com/2018/02/07/google-photos-can-now-automatically-create-themed-movies-on-demand/>, 2018. Accessed: 2018-07-25.
- [174] Ars Technica. Only a small percentage of users buys stuff through alexa, report claims. <https://arstechnica.com/gadgets/2018/08/only-a-small-percentage-of-users-buy-stuff-through-alexa-report-claims/>, 2018. Accessed: 2018-08-27.
- [175] Pui-Lai To, Chechen Liao, and Tzu-Hua Lin. Shopping motivations on internet: A study based on utilitarian and hedonic value. *Technovation*, 27(12):774–787, 2007.
- [176] Pui-Lai To and E-Ping Sung. Internet shopping: a study based on hedonic value and flow theory. *World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, 9(7):2221–2224, 2015.

- [177] Scott R Turner. *Minstrel: A computer model of creativity and storytelling*. 1994.
- [178] Scott R Turner. *The creative process: A computer model of storytelling and creativity*. Psychology Press, 2014.
- [179] The Verge. Spotify and genius are collaborating on info-rich behind the lyrics playlists. <https://www.theverge.com/2016/1/12/10750990/spotify-genius-behind-the-lyrics-playlists-iphone>, 2016. Accessed: 2018-07-25.
- [180] Bonnie C Wade. *Thinking musically: Experiencing music, expressing culture*. Oxford University Press New York, 2004.
- [181] Robert A Westbrook and William C Black. A motivation-based shopper typology. *Journal of retailing*, 1985.
- [182] Robert Wilensky. Story grammars versus story points. *Behavioral and Brain Sciences*, 6(04):579–591, 1983.
- [183] R Michael Young, Stephen G Ware, Brad A Cassell, and Justus Robertson. Plans and planning in narrative generation: a review of plan-based approaches to the generation of story, discourse and interactivity in narratives. *Sprache und Datenverarbeitung, Special Issue on Formal and Computational Models of Narrative*, 37(1-2):41–64, 2013.
- [184] Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *Proceedings of the IEEE international conference on computer vision*, pages 19–27, 2015.
- [185] Vilmos Zsombori, Michael Frantzis, Rodrigo Laiola Guimaraes, Marian Florin Ursu, Pablo Cesar, Ian Kegel, Roland Craigie, and Dick CA Bulterman. Automatic generation of video narratives from shared ugc. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, pages 325–334. ACM, 2011.