ABSTRACT

AZAD, SASHA. The Little Computer People Taxonomy and Social Physics Engine. (Under the direction of Chris Martens and Arnav Jhala).

The believability of social characters and their impact on the virtual worlds they belong to has long been studied and explored in art and literature (Thomas et al. 1981). Modeling social agents is widely considered to be a long-term research goal in the fields of entertainment (Togelius et al. 2013; Yannakakis 2012), ecology (Grimm and Railsback 2005), and computational social science (Siebers et al. 2010). Bolstered by this growing interest in simulating social agents, research on agent models has spanned topics such as planning, procedural storytelling, and decision-making.

However, social simulation research groups work in isolation, designing and discussing their character models with disparate approaches, often using project-specific terminology. This disparity makes identifying, classifying, and accumulating existing knowledge in the domain challenging. I propose that since the modeling of agents has become an integral part of the scientific practice in our field, we must develop a common taxonomy and framework to discuss and simulate these agents. With this goal in mind, we conducted an in-depth analysis of a selection of projects, categorizing existing agent social interactions and comparing results from research-based and commercial social simulation works in the entertainment domain. We conceptualized a taxonomy that classifies agent interactions by their inter-agent behaviors, formalizes rules for the interaction space, and describes the effect of these behaviors on the social state of the agents. Further, I was able to combine methodologies from agent-based simulation systems and discrete event simulation to create a modular and flexible social physics engine. To our knowledge, this is the first time these methodologies have been integrated in this way.

This proposal describes our efforts to evaluate and operationalize a taxonomy and an accompanying framework or “social physics engine”. Our taxonomy and accompanying framework would allow scientists to reproduce and evaluate existing models, collaborate on standards, share advances with other researchers and practitioners, allow for better communication and methodologies developed for new techniques, and allow for a more rigorous model-to-model analysis.

The Little Computer People Taxonomy and Social Physics Engine

by
Sasha Azad

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Computer Science

Raleigh, North Carolina
2024

APPROVED BY:

Chris Martens Co-chair of Advisory Committee	Arnav Jhala Co-chair of Advisory Committee
James Lester	Adriana de Souza e Silva

Gillian Smith

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

> *Characters whose adventures and misfortunes make people laugh and cry... that appear to think, make decisions, and act of their own volition. It is what creates the illusion of life (Thomas et al. 1981).*

Human beings are social and complex. The modeling of social characters and their impact on the virtual worlds they belong to has long been studied and explored in art and literature (Thomas et al. 1981). Yannakakis (2012) posited that there are four big areas of research within the Game AI field: player experience modeling, procedural content generation (PCG), data mining of game information, and improving social character behaviours and capabilities. This proposal tackles the latter. Creating believable simulations of human behaviour in virtual worlds via social agents is widely considered to be a long-term research goal in the fields of entertainment (Togelius et al. 2013; Yannakakis 2012), ecology (Grimm and Railsback 2005), and computational social science (Siebers et al. 2010).

In pursuit of this goal, the industry has developed virtual social characters for games franchises such as *Stardew Valley* (LTD 2017), *The Sims* (Maxis 2003), and *Animal Crossing* (Nintendo 2001). These games achieve consistent popularity with diverse audiences of players who report that interacting with and building relationships with virtual characters positively impacts their play experience (Thomas et al. 1981; Guo 2020; Lyon 2020; Hernandez 2015). Revenue in the digital games industry has grown from $69 billion in 2015 to just over $347 billion in 2022.

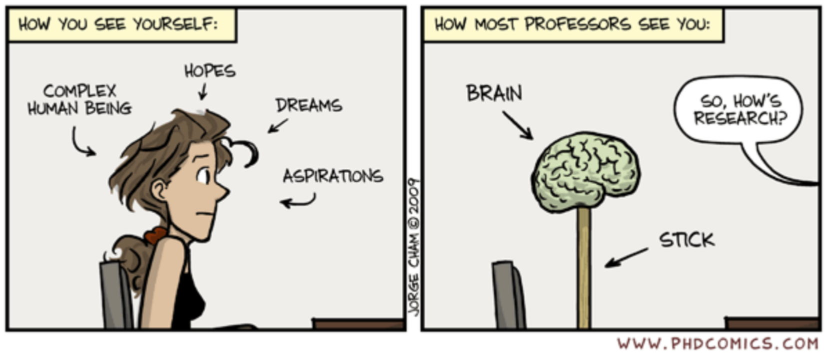


Figure 1.1: *Brain on a stick (Cham 2009)*. Human beings are complex.

The research goal of simulating social interactions among virtual characters has a large impact not only in the entertainment industry but also in improving humanity's predictive power about human social behavior. Social intelligence research groups aim to advance the ability of virtual characters to have believable interactions across a wide variety of applications ranging from entertainment to anti-bullying interventions (Pynadath and Marsella 2005), pedagogical use cases for cultural skills (Morrison and Martens 2018), or even military decision-making training (Johnson et al. 2004; Zook et al. 2012). Further, there is an increasing interest in the phenomenon of *emergence* in social simulation. Emergence occurs at the system level: system behaviors or narratives may emerge from the traits of individuals as the individuals interact with each other. Emergent behaviors can also include patterns of individual behavior that arise within the system: such patterns are properties of the entire system (Grimm and Railsback 2005). For instance, analyzing traffic patterns during a work week or studying group dynamics and shifting relationships among conversationalists during altercations (Dickinson et al. 2017; Azad and Martens 2019).

Social intelligence research aims to advance the ability of virtual characters to make believable decisions in social situations, such as conversation. This work has arisen in contexts including procedural storytelling (Meehan 1977; Ryan et al. 2015; Mateas and Stern 2003), narrative planning (Young et al. 2013; Eger 2020), knowledge and belief propagation (Azad and Martens 2019; Ryan et al. 2015), social relationship dynamics (Ryan 2016;

Electronic Arts 2000; Ryan et al. 2016a), group behavior (Azad and Martens 2019), and more. Collectively, this set of approaches for computationally enacting inter-character behavior, intended to model some aspects of human interaction, is known as *social simulation*.

1.1 Motivation

Despite the plethora of social character models developed across applications, there is *little to no agreement* on what constitutes social simulation. For some researchers and developers, it means simulating human behavior on a timescale of hundreds of years, encompassing the rise and fall of civilizations and societies (Ryan et al. 2015; Adams and Adams 2006). At this scale, characters are more or less interchangeable because it is impossible to examine each character at an acceptable level of detail. For others, social simulation refers to the minute intricacies of conversation between a small, fixed cast of characters, whose lines of dialogue and range of possible behaviors are manually authored and differ for each specific character (Mateas and Stern 2003; Morrison and Martens 2018; Evans and Short 2014). Still, others model the social dynamics of interchangeable sets of characters who can form and break alliances or trust (McCoy et al. 2011a; Azad and Martens 2019).

Instead, social simulation research groups tend to work in isolation, creating new simulations from scratch without building on prior work, discussing their advances with disparate approaches, and often using project-specific terminology. These disparate approaches result in a proliferation of disparate implementations. Researchers either (1) use similar terminology to describe either wildly different social phenomena, (2) use conflicting terminology to describe the same phenomenon, or (3) agree on terminology but model the underlying social phenomenon with varying levels of computational granularity. For instance, several research projects describe "social relationships" amongst characters as an integral part of their design. However, while one project models "social relationships" as an edge in a social network graph with a valence depicting its strength (Azad and Martens 2019; Ryan et al. 2015), another may associate complex dimensions of trust (Meehan 1977), or social rank (McCoy et al. 2011a). This lack of consensus as well as the inconsistencies in the vocabulary describing research advances in the design of social character models, make it challenging to identify, classify, accumulate, or build on existing knowledge in the social simulation domain.

Further, this lack of consensus has several disadvantages and pain points associated with it. First, it impairs *reuse*, including the ability for outside groups and new members of the field to contribute by building on prior work. Most groups develop their simulations from scratch without building on prior work, resulting in a proliferation of disparate implementations using similar words to describe wildly different social and computational phenomena. Second, it impairs *reproducibility* of research: system descriptions and results of experiments tend to be reported in ways that rely too much on metaphor and intuitive explanation rather than formal definition or open-source code, making it challenging for others to reproduce (Hales et al. 2003; Richiardi et al. 2006). Finally, these two problems together impair *comparability*: we cannot meaningfully measure or even qualitatively compare the effects of two systems with superficially similar sets of simulated social phenomena.

In other game authoring technologies, such as physics engines or graphical rendering systems, the standardization of features has led to these technologies being widely available in commercial and hobbyist game development platforms. We envision a future in which the same can be said for social simulation via a "social physics engine," to borrow a term from McCoy et al. (2010b). To do this, we need to identify and formalize a common language of social interactions that can be used to tell many different kinds of stories, given at a level of specificity that it can be implemented as an API or library in any game development environment.

1.2 Goals

My proposed research identifies and addresses the disadvantages caused by the use of disparate methodologies and the identified lack of consensus. I argue that the narrative and entertainment AI community needs conceptual tools that help researchers map the design space for social character behaviors, comparing individual behaviors and considering how the behaviors could interact. Furthermore, researchers need to be able to reproduce and compare different social simulation systems.

Towards this end, I detail my three core research goals and describe the design and development of their associated research artifacts that will help achieve my goals to support social simulation research.

My first goal is to eliminate the inconsistencies in the vocabulary currently used to describe and evaluate existing models. I hypothesize that eliminating inconsistencies in the vocabulary and consolidating the differences and similarities observed across social simulations would enable researchers to better understand the underlying social theories described. Towards this goal, my first research artifact is a taxonomy that is designed through an in-depth analysis of a selection of projects and maps current advances in the field of social simulation. With my preliminary work, I was able to identify and formalize a common lexicon and rule set, the *Little Computer People Taxonomy* (Azad and Martens 2021).

My second goal is to aid researchers and practitioners by developing a *standardized* framework for modeling social characters. Game developers are typically used to standardized physics engines. Such standardization of features has led to the democratization of these technologies, making them widely available in commercial and hobbyist game development platforms. I envision a future in which the same can be said for social simulation via a "social physics engine" (McCoy et al. 2010b). Thus, my second artefact is the proposed design and development of a social physics framework that allows researchers to model and design social simulations. This framework will be tested on two case studies, Lyra (Azad and Martens 2019, 2018) and Anthology (Azad et al. 2022), to study the ease of incorporating the designed modular framework into existing projects. These case studies will further help in the requirement engineering phase of the design of the framework.

In particular, I focus on how my contribution will aid two specific user groups in their goals:

- Experienced Researchers — as *producers of social simulations*, with their goals to design and research social simulations, more easily reproduce and evaluate existing models, and collaborate on standards with other researchers or industry practitioners,

- Game Designers or Computational Social Scientists — *as consumers of social simulations*, by helping them understand existing research and the underlying social theories modeled, and easing their barrier of entry.

Finally, I will evaluate my proposed taxonomy and framework focused on understanding the extent to which these two artefacts achieve my goals and support my identified user groups.

My dissertation aims to help social simulation researchers and practitioners in their goals to understand the depth and breadth of existing research and author new social simulation theories, as well as further the design of social character authoring tools that are beneficial for the narrative and entertainment interaction community. I believe my work lays the groundwork for robust future social interaction research.

1.3 Pain Points: The Need for a Taxonomy

Our taxonomy-building work arises from observing several pain points in our research and implementation practice. Across social simulation artefacts, we come across terms such as “social state, relationships, influence, etc.” However, missing a formal definition, these terms can mean many things. For instance, the term social state refers to any combination of emotions, relationships, personality or intentions.

Furthermore, this lack of consensus impairs reuse, including the ability for outside groups or new members of the field to contribute by building on prior work. Second, it impairs reproducibility: with system descriptions and results of experiments reported in ways that rely too much on metaphor and intuitive explanation rather than formal definition. Next, these two problems together impair comparability, making it difficult, if not impossible, to compare the effects of two systems or features meaningfully. Finally, we hope this will ease the way to better research collaboration.

1.4 Thesis Statement and Research Questions

The main goal of my dissertation is to design a social character authoring framework that allows researchers and practitioners from the domain to better understand and map existing research advances, allowing them to reproduce and evaluate existing models and develop new models easily. The research questions described above lead to my overall thesis statement:

When social simulation researchers and practitioners use tools such as a common vocabulary and social character authoring framework, they will be able to better understand and contextualize new and existing research advances, create computer simulations that better match their mental models of underlying social phenomena, improve reuse and reproduction of published models, and more meaningfully evaluate and compare social simulation research efforts.

My work has five research thrusts, each with its own set of research questions or goals. Together the research questions serve to answer my overarching thesis statement (see section 1.4) describing how we could design social character authoring tools that are beneficial for the social simulation community laying the groundwork for robust future social interaction research.

Thrust #1 Survey of Existing Research and Commercial Social Simulations (Completed)

Thrust #2 Designing the Taxonomy (Completed)

Thrust #3 Developing Social Simulation Case Studies (Completed)

Thrust #4 Developing a Social Simulation Framework (In Progress)

Thrust #5 Evaluating the Impact (In progress)

These research thrusts lay the foundation for my dissertation work. Furthermore, they are (partially) chronological [1]. Finally, these thrusts divide my work into logical segments, each of which with associated research questions and sub-questions. Finally, these five research thrusts describe the structure and organization of this proposal. The research I propose is focused on how the Little Computer People taxonomy and its accompanying framework can be used by experienced social simulation researchers, game developers, or the users of social simulations to design, develop, evaluate, or analyze both previously deployed and new social simulations.

1.4.1 Thrust #1: Survey of Existing Social Simulations

The first research thrust *(completed)* addresses my first research question (RQ1). The first thrust of our project was to understand and organize knowledge concerning existing design processes and methodologies involved in the development of social simulations.

***RQ 1:** How do multi-agent social simulations in the narrative intelligence domain currently model social inter-agent behaviors?*

To answer RQ1, I conducted an in-depth analysis of a selection of research projects and commercially available popular social simulation game franchises. I collated existing knowledge concerning agents modelled by these simulations. Furthermore, I was able to identify a lack of consensus and inconsistencies in the vocabulary used to define and

[1] The exception to the chronological order is that that the design and development of our first case-study (from Thrust #3), Lyra (Azad and Martens 2019), occurred simultaneously alongside Thrust #1 and motivated the design of the taxonomy in Thrust #2.

describe research advances. This lack of consensus directed my inquiry and analysis of existing research projects.

In chapter 4, I discuss the methodology and findings of the survey and the inconsistencies in vocabulary discovered.

1.4.2 Thrust #2: Designing The Taxonomy

The second research thrust *(completed)* addresses my second research question (RQ2).

> ***RQ 2:*** *How can we consolidate the differences and similarities currently modelled in the social simulation agents?*

To resolve the discovered inconsistencies, I organized the knowledge collected for RQ1 to design a common lexicon and rule set for modeling social agents.
Research Thrust #2 resulted in the design and development of The Little Computer People taxonomy (LCP) (Azad and Martens 2021). The vocabulary and nomenclature decisions in the construction of the taxonomy were made by consulting existing vocabularies, project publications, social science publications, or other existing logic and computer science terminologies. Further details on the process of development of the taxonomy have been detailed in Chapter 5.

1.4.3 Thrust #3: Designing Social Simulation Case Studies

For my third research thrust *(completed)*, we anticipated early users of our framework to be one of the following three typical use cases .

- Use Case #1: Early users may simulate social characters as a means to produce new emergent or interactive narrative experiences (Mateas and Stern 2003; McCoy et al. 2012), or
- Use Case #2: Early users may use our social simulation framework as a means to examine or evaluate underlying decision-making or computational social science theories in existing simulations.
- Use Case #3: Early users may use our framework as a means to design and augment existing social character simulation models.

We anticipated early users of our taxonomy and framework to have one of the use cases highlighted above. I designed two accompanying social simulations, Lyra (Azad and Martens 2019, 2018) and Anthology (Azad et al. 2022), to serve as case studies to test the use cases outlined above. These simulations will aid in the requirement-gathering phase while developing the framework. Chapters 3 and 6 will describe my two simulations in further detail. We will incorporate the final LCP framework into the social simulations. Finally, I plan to test the integration of the final frame with the Lyra and Anthology simulations to understand the ease and impact of incorporating the designed modular framework into them.

1.4.4 Thrust #4: Developing the Framework

With Research Thrust #3 *(proposed)* I plan to generate a social character framework or authoring library that developers can use to design and develop social agents. I propose designing and developing my second research artefact, a "Social Physics Framework" (McCoy et al. 2011a). The development of this tool would address the question:

> ***RQ 3:*** *How can we operationalize the designed taxonomy into a framework that our identified user groups can use?*

We will be using the lessons we learned and reusing successful chunks of the simulation code from our previous case studies (described in Chapters 3 and 6) to form the base of the simulation.

The Social Physics Framework would be a significant software engineering effort. I anticipate that the initial informal feedback I obtain from users will help direct the design choices of the project. The feedback will also help with the development of the framework to the extent that the initial release of the social simulation authoring tool is usable and friendly. I believe the development of this framework would be a valuable research artefact contributed by my work.

Finally, I will describe how early versions of my proposed framework are being used within the context of an enterprise-deployed application of the digital twin paradigm (Tao et al. 2018). As part of a summer internship, I designed and developed a locative digital twin social simulation for employees inhabiting and interacting within a workplace. The simulation was aimed at understanding the associated risks and spread of COVID-19 within the workplace. My simulation took into account location-specific resource bottlenecks, allowed executive management to test various health policies and strategies, and analyzed their effect on the spread of the virus whilst remaining privacy-preserving. This final simulation will serve as a way to analyze how merging the social and cognitive models from Agent-Based Social Simulation (ABSS) with methodologies from Discrete Event Simulation (DES) could add value to deployed applications.

1.4.5 Thrust #5: Evaluating the Impact

Finally, for my fifth thrust *(proposed)*, I propose to evaluate my framework's impact on novice and experienced researchers in terms of its applicability to their work and goals. Toward this end, I have outlined the research question associated with my evaluation.

RQ 4: *What is the impact of the taxonomy and framework on both experienced social simulation researchers in terms of its applicability to their modeling process? Can the taxonomy and framework be used to evaluate existing social simulations by users of the simulations?*

It is not necessary to complete the framework for the evaluation of the taxonomy to begin. We plan to start the human subject study, and summative evaluation in parallel, using the taxonomy to form the basis of our questions (described in Section 8.3). We have already obtained some informal feedback from social simulation researchers (see Section 7.1) on our preliminary work during conference presentations of our publications (Azad and Martens 2021; Azad et al. 2022) on the same.

CHAPTER

2

BACKGROUND AND RELATED WORK

In this chapter, we identify the prior work done in social dynamics and psychology to situate the need for similar research and analysis for virtual characters or non-player characters (NPCs). Next, we describe related work from the narrative domain on believable virtual characters. We further introduce the reader to a class of virtual characters, called social characters or social agents. Finally, we discuss key existing shared vocabularies and taxonomies that have helped guide our taxonomy design and methodology.

2.1 Social Dynamics and Pscyhology

Social Dynamics (Durlauf and Young 2001) is the explicit study of the interactions linking individual behavior and group outcomes. Social psychology (Stangor 2020) is the study of the dynamic relationship between individuals and the people around them. In their work, Allport (1968) described why social behaviors were necessary to be studied, stating, "Social psychology attempts to understand and explain how the thoughts, feelings, and behaviors are influenced by actual, imagined, or implied presence of others" (Allport 1968). Social behaviors are emergent from the formation of relationships, groups, and institutions among individuals, independent of individual characteristics.

Lewin (1951) developed a theory that emphasized the importance of individual personalities, interpersonal conflict, and situational variables. He formalized this joint influence of an individual's characteristics and the situational variables with the following equation (Lewin 1951):

$$behavior = f(individual\ characteristics, social\ situation) \tag{2.1}$$

Lewin (1951) argued that behavior should be defined as a function of both the individual characteristics or personality of the agent and the environment or social situation in which the agent is functioning (as shown in Eq. 2.1), with the added complications that environment is a function of personality, and personality a function of environment.

With our work, we propose a classification of existing behaviors. Our goal is to inform novice developers of existing models of social interactions and features that can be varied or result from their library of individual characteristics or social situations or phenomena they choose to simulate in their work.

2.2 Impact of Believability of Social Characters

There is no generally agreed-upon definition of believability. Instead, within the narrative field, believability is used linguistically to describe that which is believable by someone. In terms of virtual characters, this could imply some aspect of their viewed interactions (either with the player or with each other) is believable.

Hayes-Roth and Doyle (1998) claimed that “animated characters” redefined traditional agent design problems. Agent behaviors must be variable rather than reliable, idiosyncratic instead of predictable, appropriate rather than correct, effective instead of complete, interesting rather than efficient, and distinctively individual instead of optimal. Togelius et al. (2013) describe how games that incorporate believable elements can elicit particular emotional responses to a player. They discuss how the generation of believable, human-like opponents leads to increased player enjoyment. Additionally, rich social interactions among NPCs have been found to improve the believability of interactive narratives and the player experience (Afonso and Prada 2008; Swartout et al. 2006). Improving the believability of interactions is very important because it transforms the challenge of an experience from a being technical one to an interpersonal one, increasing both the enjoyment and the engagement of players.

Furthermore, there is no agreed-upon evaluation of believability in agents. Bates et al. (1994) claimed that the most significant requirement for believability in computer characters is appropriately timed and clearly expressed emotions that make characters seem more life-like. However, emotion alone is insufficient. Thus, prior research has evaluated how factors such as cognition, decision-making, desires, motivations, goal-based behavior, and more have an effect on the believability of agents.

2.3 Social Agents

The entertainment industry has long studied virtual characters. Prior work (Wooldridge and Jennings 1994; Moulin and Brahim 1996; Brassel et al. 1997) had broadly categorized agent models into *Reactive agents*, *Intentional agents* and *Social Agents*. Reactive agents receive environment stimuli and react to the same using fixed rules, without any reasoning, planning, or undertaking deliberative actions (Franklin and Graesser 1997). Intentional agents include meta-rules to define goals and are capable of detecting goal conflict within specified bounds. Social agents contain models of other agents and themselves and are capable of utilizing these internal models to reason about their goals, expectations, and motives and incorporate these into their actions.

This dissertation focuses on social agents. Social agents are defined as being cognitive (Kugele and Franklin 2013) in nature. These agents are capable of utilizing internal models and thought processes that are detached from current sensory stimuli as cognitive agents. Rather than focusing purely on *online cognitive modeling* of processes, that react to the environment stimuli received, our research focuses on simulating the cognitive and social processes amongst agents or agents that are capable of *offline cognitive modeling* processes (Kugele and Franklin 2013). These processes enable the construction (and manipulation) of imagined realities. Planning, reasoning, introspection, and problem-solving, as well as more pedestrian activities, like the recall of long-term memories and daydreaming, are all “offline” cognitive processes.

Extensive research has been conducted on social rules and interactions between social agents. Versu (Evans and Short 2014) shows characters interacting with one another using pre-constructed social practices templates. These templates are constructed manually, can be time-intensive, and require domain knowledge. Similarly with CiF, in Prom Week (McCoy et al. 2011a), the authors describe a social physics architecture model that constrains how NPCs behave. Social simulations are defined as a series of multi-character social interactions. Each interaction modifies the social state existing within the agent itself and across any other participating agents. McCoy et al. (2010a) describe their social games as being heavily influenced by Goffman's dramaturgical analysis (Goffman et al. 2002) as a way to encode normal patterns of behavior in terms of how a character would present themselves in order to manage how they are perceived by others.

With regards to the social relationships formed between agents, the network and interactions amongst the actors takes shape based on their relationships with one another. The amalgamation of the entities in the network is an abstraction that is necessary only analytically. For instance, simulating a "city government" involves simulating a messy combination of agencies, elected officials, press releases, etc. However, it is the ways these components interact that we mean when we use the term city government (Wessells 2007). Similarly, with the relationships, "employee-manager" or "student-teacher", these terms allow us to analyze the types of interactions that can be simulated between the entities participating in the relationship.

Brassel et al. (1997) described how describing a simulated social agent model usually consist of the following details:

- *Agents:* A group of one or more agents representing the individuals that make up the real world or system being emulated. Stangor (2020) describes how the individuality of the agents or subjects must be maintained at all times. An individual's personality traits, desires, motivations, and emotions have an important impact on their social behavior.

- *Behavior or Interaction Modes:* A subset of distributed, discrete and localized behaviors that encapsulate the phenomenon, operations, or interactions being studied. There should be flexibility to incorporate different kinds of agent models to represent real-world agents. It should support the design of heterogeneous multi-agent systems, i.e., systems with agents differing in behavior and capabilities such as communication, movement, etc.

- *Representation of the environment:* Conceptual and technical representations of the environment in which the agents interact, further categorized as a:

  – *A Common Environment* to be one common to all agents in the simulation or at a given level of aggregation.

  – *A Specific Environment* for an agent to be comparably complex; or for agents (or environment) to interact specifically with it (can change over time).

Our work focuses on simulations of large populations of social agents (Wooldridge and Jennings 1994; Moulin and Brahim 1996; Brassel et al. 1997). We conceptualize a taxonomy to better communicate the *social behaviors or interaction modes* (Brassel et al. 1997; Hayes-Roth and Doyle 1998) described above.

2.4 Existing Social Simulations or Architectures

We overview some of the available social simulation architectures and discuss how they differ from ours. A more in-depth survey of a selection of social simulations has been detailed in Chapter 4.

PsychSim (Pynadath and Marsella 2005) is a simulations framework that produces social interactions between multiple agents in an environment. PsychSim agents have their own decision-theoretic model of the world, including beliefs about their environment and recursive models of other agents (Pynadath and Marsella 2005). For instance, a bully may believe that their teacher is weak. Agents in PsychSim have social relationships with other agents, as well as preferences about the world. For instance, an agent may have a preference to maximize laughter generated by any onlookers. The simulation is run, behaviors are invoked based on the scenario, environment, and beliefs, and the system generates explanations for the agent behaviors.

FAtiMA (Dias and Paiva 2005) takes into account emotion, appraisal theory, and agent goals while modeling their agents. The system uses a computational theory of emotion and appraisal modeling termed the OCC (Ortony et al. 2022). FAtiMA agents are able to simulate empathy and culture in interactions with one another. Agent cognitive parameters can also be adjusted, allowing for varying utility evaluation while comparing possible behaviors to undertake. For instance, events producing states of high arousal of emotions require a more urgent “call-to-action” and thus a faster decision process requiring tradeoffs.

Comme il Faut (CiF) (McCoy et al. 2010a,b) is a social simulation framework and authoring tool for creating game-based interactive stories about relationships and social interactions amongst characters. CiF characters take into account their interaction history with one another and a set of social rules depicting good or normal social behaviors before choosing which behaviors can fire. CiF is best known as being the underlying architecture behind the game, PromWeek (McCoy et al. 2011a). Players in PromWeek are assigned goals, for instance, becoming prom queen, or ask out the popular boy, etc. Players may undertake interactions to fulfill these goals, updating and affecting their social status and the social state of the characters around them.

Versu (Evans and Short 2014) is an interactive drama simulation. The system describes the concept of "social practices" that dictate the choices of the AI characters in the system. A social practice describes a social situation filled with individual agents' comments, looks, or actions. The agents use a utility-based, reactive planning algorithm to choose their next actions. Actions are written to be agnostic to agents but specific to roles. Versu agents are autonomous agents with beliefs about the world, social relationships, and basic emotions.

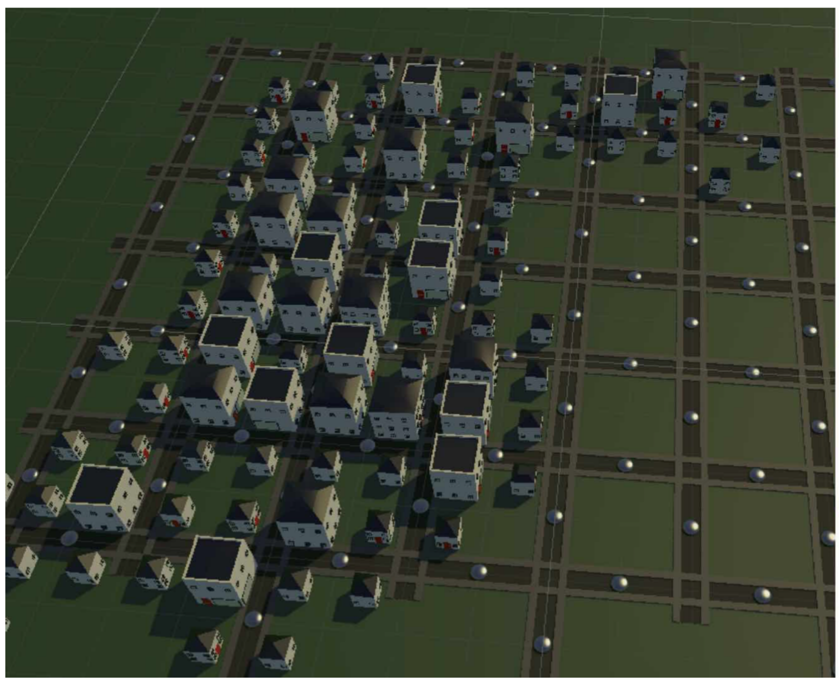


Figure 2.1: Unity prototype of an early Talk of the Town simulation (Ryan 2018)

Talk of the Town (Ryan et al. 2015) simulates a small American town with a few hundred characters that populate the same. A few days of each year are simulated (with two-time steps each – a day step and a night step), and the rest of the year is extrapolated from the same based on how long it's been since the last time step was modeled. Characters have relationships with one another that are familial, friendly, employment-based, or romantic that are governed by their initial "spark" for each other (signaling, for instance, the start of their romantic attachment) and the subsequent charging of the same (based on how frequently they interact, the salience of their relationship, and the nature of the interactions).

In a short-term project where we collaborated with a small independent game studio, *Blast Bit*, we describe our experiences with integrating procedural content generation, decision-making methods, and game artificial intelligence methodologies to generate a dwarf-like (Adams and Adams 2006) world with biomes, characters that inhabit the world, and NPC quests and interactions that can be generated for the world. This resulted in publications to the 2021 Experimental AI in Games Workshop in the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (Lech et al. 2021; Jonasson et al. 2021).

2.5 An Overview of Modeling Philosophies

With this section, we overview the current existing simulation and modeling methodologies used in Agent-Based Social Simulation Modeling and strategies from Discrete Event Simulation Modeling. We compare and contrast the two methodologies and their relationship to *emergent* and *imposed* behaviors. These differences have been summarized in Table 2.1 above.

We discuss the shortcomings and advantages we could gain using our proposed hybrid approach combining both strategies. My proposed framework would integrate ABSS NPCs with believable interactions and individual intelligence with simulation strengths from Discrete Event Simulation (DES). Thus, agents would maintain their decision-making, personalities, or control in the system using traditional NPC simulation techniques used in my surveyed simulation artifacts (eg. using planning, reactive planning, goal-based simulation or behavior trees). However, once these individual agent decisions are made, they would go to a DES-based centralized queuing system to handle the simulation of events, simulation clock increments, resource and scheduling queuing bottlenecks, and overall event control. That is, agents are modeled as "active," exhibiting cognition, while systems are modeled on a macro-level using a top-down approach.

2.5.1 Agent Based Social Simulation (ABSS)

Traditionally in the computational social simulation or entertainment domains, researchers have focused on heterogeneous human populations, modeled using a combination of Artificial Intelligence or Cognitive Science algorithms with social simulation or agent-based models. These strategies together are termed as Agent-Based Social Simulation methodologies (Davidsson 2002).

Table 2.1: Summary of the architectural differences between Agent-Based Simulation and Discrete Event Simulation models (Siebers et al. 2010)

DES models	**ABS models**
Process-oriented (top-down modelling approach); the focus is on modelling the system in detail, not the entities	Individual-based (bottom-up modelling approach); the focus is on modelling the entities and interactions between them
Top-down modelling approach	Bottom-up modelling approach
One thread of control (centralized)	Each agent has its own thread of control (decentralized)
Passive entities; Intelligence (eg, decision making) is centralized, modelled as part of the system	Active entities, that is the entities themselves can take on the initiative to do something; intelligence is represented within each individual entity
Queues are a key element	No explicit accommodation for resource bottlenecks or queues
Flow of entities through a system; macro behavior is modelled	No concept of flows; macro behavior is not modelled, it emerges from the micro-decisions of the individual agents
Input distributions are often based on collected/measured (objective) data	Input distributions are often based on theories or subjective data

In Agent-Based Social Simulation (ABSS) methodologies, the simulation takes on a bottom-up approach. Systems are made up of a myriad of agents, classified as *active entities* (Siebers et al. 2010), which indicate agents have their own behavior and their own decision-making algorithms. Designers of ABSS use computational social science theories to model complex social agents, with rich cognitive models and inter-agent communication, abstracting relationships (Ryan et al. 2016b), social psychology (Mateas and Stern 2003), emotion models (Dias et al. 2014), social behaviors (Azad and Martens 2019; Mateas and Stern 2003; McCoy et al. 2012), cultural norms (McCoy et al. 2011a; Azad and Martens 2019) and more.

ABSS is considered to be a 'bottom-up' modeling paradigm in which system-level patterns (or macro behaviors) emerge or are discovered by analyzing the behavior of individual local-level agent interactions (or at the micro-level). These observed patterns are then used to determine and analyze properties of the actual social system (Davidsson 2002; **?**). ABSS has helped us better model and understand real-world systems or behavior where modeling each individual with autonomous decision-making is important. Agent-based models can explicitly model the complexity arising from individual actions and interactions that arise in the real world (Siebers et al. 2010). This is because the diversity of human behaviors is more accurately depicted by the use of simulation (Majid and Herawan 2013; Azad and Martens 2021). An in-depth survey on MABS modeling within the entertainment domain has been described in the work, *The Little Computer People Taxonomy* (Azad and Martens 2021).

The phenomenon simulated in ABSS systems is the concept of an event. An event may be an agent interaction, an introspective event, a change in state for the agent or the system, etc. These events can be scheduled or added to the simulation on an ad-hoc basis. Instead of simply enforcing specific events to exist, with ABSS, we model the underlying, individual-level mechanisms that give rise to the behaviors. *Emergence* occurs at the system level: system-wide patterns or behaviors may emerge from the traits of individuals as the individuals interact with each other. Different emergent properties may arise when the same individuals are placed in a different environments. Emergent properties and patterns are not the sum of the properties of the individual and cannot immediately be predicted by looking solely at an individual.

Typically ABSS systems do not handle event queuing. Instead, individual agents line up for a task, and the developer is left to construct a resource or location bottleneck that handles queuing. The agent behaviors are stochastic in nature and can use some probability distributions or random probabilities to model individual characteristics to aid in decision-making (Ryan 2018).

However, most of the ABSS methodologies are computationally heavy for detailed simulations. Behavior trees or Planning methods typically use incremental time progression for the simulation and involve iterating through each agent, carrying out extensive computations to select the next interaction, or iterating through every agent's tree to check the state of the agent's current action and plan the next move. Thus, every tick of a simulation clock is incremented by a fixed duration (either resembling real-time or a period of simulation time). Additionally, there is the burden of simulating individual decision-making while coordinating among agents to work on plans or actions concurrently.

2.5.2 Discrete Event Simulation (DES)

In contrast, Discrete Event Simulations (DES) are generally used for network simulation, diagnosing process issues, operations research, and systems research. With DES, we can map interactions as discrete events independent from one another. Time in such a system need not flow incrementally. Instead, events can be decomposed into a set of logically separate processes that autonomously progress through time. Each event informs the DES engine of a specific tick at which it will be run. On completion, events can generate results or can spawn future events to be added to the timeline accordingly.

However, DES has its own flaws; agents are "passive entities" (Maidstone 2012; Siebers et al. 2010), instead of individual decisions, decision-making is centralized as they move through the system; instead, intelligence is modeled as part of the system.

2.6 General and Modular Social Agent Frameworks

Agent-based social simulation (ABSS) has been shown to be an effective tool for creating dynamic emergent narratives in video games or computational social science research. ABSS model characters as individual agents with attributes that drive their social interactions/behaviors with other characters and/or the player (Johnson-Bey et al. 2022b).

Most recently, works published at the 2022 AAAI's Artificial Intelligence and Digital Entertainment (AIIDE) conference were noted to shift to a more maximalist approach. The conference had previously seen submissions of mainly niche, one-off systems designed to work for a specific purpose. The unfortunate result of this practice is that research in the field is fragmented and largely independent, making it hard to draw generalized conclusions (Azad and Martens 2021; Mori et al. 2022). In contrast, there has been a trend towards a more maximalist approach with more general macro-systems, modular enough to support a large variety of purposes, and that could be used by a wide range of audience (Ware and Eger 2022; Mori et al. 2022).

This can be seen with systems such as Mori et al. (2022) EM-Glue that include decoupling the experience manager (EM) from the environment, thus enabling the data gathering needed to gain insights into how their EM works across different contexts.

With the research done in Villanelle (Martens et al. 2018), we examined interactive narrative authoring challenges as they pertained to social characters and NPCs. We combined insights from narrative planning techniques, reactive planning, and behavior trees to form the Villanelle software project. The work resulted in a language-based social character authoring tool that can be used for generating interactive narratives.

Johnson-Bey et al. (2022b) describes Neighborly as a customizable, community-scale social simulation engine for simulating a population of characters inhabiting a small town. Neighborly is a rational reconstruction of Talk of the Town (TotT) (Ryan et al. 2015), an earlier social simulation for emergent narrative focused on simulating small American towns and the townspeople's lives. TotT has been described in more detail in Section 2.4. However, the Neighborly reconstruction, labeled TotT-like, is more modular and generalizable, allowing for more authorial control over character roles, businesses, and internal decision-making logic.

Built on the Skyrim Creation Kit and inspired by its namesake architecture, Comme il-Faut (CiF) (McCoy et al. 2010a,b) is CiF-CK (Guimaraes et al. 2017). The authors describe a stack of individual behaviors with pre-conditions. NPC behaviors are designed as scenes tied to packages. These behaviors are periodically evaluated to see when preconditions are met and the top behavior/package is run. For instance, an NPC may go home if it's midnight. Quests in Skyrim include players interacting with the resulting designed social CiF NPCs.

Most recently, as part of a summer internship with IBM Research, I designed and developed a locative digital twin social simulation for employees inhabiting and interacting within a workplace. The simulation was aimed at understanding the associated risks and spread of COVID-19 within the workplace. My simulation took into account location-specific resource bottlenecks, allowed executive management to test various health policies and strategies, and analyzed their effect on the spread of the virus whilst remaining privacy-preserving. This final simulation will serve as a way to analyze how merging the social and cognitive models from Agent-Based Social Simulation (ABSS) with methodologies from Discrete Event Simulation (DES) could add value to deployed applications. While this work is directly relevant to my dissertation and informs part of the Little Computer People Framework, I am unable to share further details about this project until the two associated research papers are published due to copyright constraints (Azad pear).

2.7 Existing Shared Vocabularies

There have been several surveys and analyses performed of existing work in the field of interactive narratives. Existing taxonomies and surveys have focused on elements such as human creativity (Gervás 2009), interactive narratives (Luo et al. 2015), computational narratives (Cavazza and Pizzi 2006) and narrative planning (Young et al. 2013).

Human behavior is unpredictable, and adding the ability for interactivity while maintaining the experience's narrative is a problem many have tried to solve. Researchers have created a spectrum of audience interaction (Striner et al. 2019), surveying ways in which audiences of interactive entertainment experiences can express their individuality and agency as consumers of interactive experiences. *A Spectrum of Audience Interactivity for Entertainment Domains* was my first foray (during the PhD) with regards to an exhaustive cross-disciplinary literature review across entertainment domains of Theme Parks, Interactive Narrative and Video Games, and Interactive Theatre. Our review evaluated, iterated, and expanded on an initially proposed vocabulary for classifying interactivity [1].

For producers of these experiences, prior work has created a vocabulary to describe, compare and interpret research in the field. In the past, shared representations of drama and experience managers (Thue and Bulitko 2018; Roberts and Isbell 2007) have evolved to characterize this need. Roberts and Isbell (2007) compared various existing drama management technologies describing a set of desiderata for the qualitative analysis of such systems. Research has also looked at varying narrative generation techniques (Roberts and Isbell 2007). Kybartas and Bidarra (2016) surveyed mixed-initiative narrative generation techniques, classifying a large body of research based on the degree of generated and manually authored content proposing a formal model of narrative to allow for collaboration in the space (Kybartas and Bidarra 2016).

With regards to social characters inhabiting these works, Tosic and Agha (2004) describe efforts to form a taxonomy of agents from a systems perspective (Tosic and Agha 2004). Lisetti (2002) describes a hierarchical taxonomy integrating personality, affect, mood, and emotion for social characters (Lisetti 2002). In another work, Zoric et al. (2007) surveyed facial gestures for embodied conversational agents to create a taxonomy and guideline for their implementation in believable agents (Zoric et al. 2007). Our work extends this research by adding a taxonomy of social interactions implemented by both scientists and practitioners of the field.

Finally, Isbister and Doyle (2002) surveyed embodied conversational agents considering several criteria such as social interface and believability. The authors described the need for a social agent to carry on discourse, understand or cooperate with other agents, have a capacity for social relationships and social behaviors (Isbister and Doyle 2002). For these behaviors and domain behaviors to be believable, they must have contextuality and display

[1] The Spectrum of Audience Interactivity was well received and obtained the Best Paper award at the 2019 International Conference on Interactive Digital Storytelling where it was presented (Striner et al. 2019)

appropriate behaviors and intentions. They should do nothing “clearly stupid or unreal,” having well-integrated capabilities with the environment (Isbister and Doyle 2002). While these prior works have been able to pinpoint the need for believable social characters with strong, coherent social behaviors, they differ from our work. This paper aims to describe a taxonomy of social interactions in virtual characters.

To the best of our knowledge, ours is the first attempt to characterize and create a taxonomy of existing social interactions, knowledge, and social relationships modeled between social agents.

CHAPTER

3

CASE STUDY: LYRA

I built the Lyra simulation to explore how to generate and simulate a social town of characters. This work was my first foray into the world of simulating social characters. I was very taken by the work done by Ryan et al. in their Talk of the Town (TotT) and Bad News project. At the time, it had been a few months since the USA 2016 elections, and all around me my friends and acquaintances at school and work were hashing out political views, and trying to change each other's minds on various topics. We were all exhausted, reading the news each day wondering what the day would bring, and then discussing the latest antics of the government and the positions of the media when we met in between classes, as well as during dinner or drinks after school. I began to wonder whether I could simulate a population that generates or simulates a similar set of arguments and discussions as I experienced every day. I imagined such contentious political discourse affecting the fabric of this society, enhancing some relationships while breaking others. Perhaps I would be able to add a voting mechanic[1], generating something along the lines of how CGP Grey's "Politics of the Animal Kingdom" showed the fluctuating of opinions and views (Grey 2011)

[1] Spoiler alert! I was a naive dreamer. I was not able to include the voting mechanic even though I designed one. It could not be included in the evaluation if I did add one because of reasons addressed in this section. I do however think it would be a very cool future side-project!

and how it could affect the vote.

That year, I started the PhD program at NC State University in 2017, and enrolled in Dr. Chris Marten's Generative Methods for Game Design class. My ideas for this politically charged generative world materialized and Lyra was conceived as the project component for the class. Lyra was a way for me to explore various research dimensions within the domain of social characters. For instance:

- How does a novice social simulation designer and engineer (as I was at the time) begin to simulate a small population of virtual characters?
- What is the required granularity of interactions simulated by the characters?
- How do we model a computational social phenomenon (in this case, political discourse and the flow of bias) through this population?
- How do we evaluate the NPCs populating the town for believability?
- How do we evaluate the computational social phenomenon we're modeling for believability?

Designing and developing Lyra project was a lot more challenging than I anticipated. I was struck by the fact that I would not be able to successfully evaluate both the social simulation and the computational social phenomenon of opinion modeling since I was modeling two complex independent variables, both of which contributed an effect on the other. For instance, if my social simulation was not believable or the granularity of the same was incorrectly modelled, this would have an adverse effect on the evaluation of my computational social opinion and bias discourse model, and vice versa.

To solve this dilemma, I decided to try to use an existing simulation and modify it to include in my political discourse component. That way, I would not need to prove my social simulation model was successful before my discourse model could be evaluated. This decision proved to be futile. Existing social simulations were too tightly enmeshed and interwoven with one another to be successfully integrated for a new application or to test my discourse model. Each simulation had a disparate approach specifically attuned to a single application. As such, I was forced to build a social simulation from scratch and Lyra was born, my first attempt to reinvent the wheel.

I anticipate early users of my taxonomy and framework to have a similar use case. Early users may (1) simulate social characters as a means to produce emergent or interactive narrative experiences (Mateas and Stern 2003; McCoy et al. 2012), or (2) use a social simulation as a means to examine underlying decision-making or computational social science theories (Azad and Martens 2019, 2018). This chapter details my first social simulation and the accompanying user study and evaluation. With this social simulation, I aim to show how much effort and background theory on social simulation is required for a novice to test a single computation social phenomenon. In Section 3.13 I discuss some of the difficulties we faced in both the construction of Lyra as well as the evaluation of the system due to the lack of an existing underlying social physics framework, and the need we found to reinvent the wheel.

3.1 An Introduction to Lyra

> Atticus: *"You never really understand a person until you consider things from his point of view–"*
> Scout: *"Sir?"*
> Atticus: *"–until you climb inside of his skin and walk around in it."*

Humans are rational and emotional beings. Their social systems are complex and contextual. The quote above from Lee's fictional yet well-beloved lawyer, Atticus Finch, is a familiar one (Lee 1960). Atticus believes that the prevalent bias and anti-social behaviour against members of their town is a consequence of not understanding different perspectives, leading to townsfolk being discriminated against or shunned. Understanding such behaviour and simulating it with virtual characters requires reasoning not just about observable social network graphs or social interactions, but also about geography, economics, and increasingly, online participation and discourse. Riedl (2016) describes machine enculturation as the act of instilling social norms, values and etiquette into computers so that they more readily relate to us, and avoid harming us. When instilling these norms into virtual characters by applying artificial intelligence, *social intelligence* is a critical form of reasoning. Wang et al. (2007) discuss how the move to *social intelligence* can be achieved by modelling and analyzing social behaviour, by capturing human social dynamics and creating artificial social agents that generate and manage actionable social knowledge.

Models to simulate such social intelligence with artificial intelligence have been used in the past to create social training environments (Morrison and Martens 2018; Fowler and Pusch 2010). In digital games with large populations of autonomous non-player characters

(NPCs), players find interactions between characters to be more believable if they adhere to recognizable social practices and plausible enculturated (Riedl and Harrison 2016) responses to social situations (Warpefelt 2016). However, these simulated models typically do not account for some of the most important features of social networks, namely that of the *social dynamics of opinion change* and its cause and effect relationship with social relationships.

One key part of social interaction is the dynamics of opinion change and its cause and effect relationship with social relationships. This form of interaction among humans has recently captured the interest of the public with our increasing understanding of the feedback loops created by social networks and political influence (Brichacek 2016). While one approach to studying this phenomenon could be to analyze data generated by real user interactions on social networks, we posit that modelling and simulation based on cognitive and social theories can produce good explanatory results of the mechanisms at play during the sharing and swaying of opinions. Correspondingly, we argue that the simulation of opinion change and the causes and effects of bias will positively affect the believability of virtual characters.

This project investigates how to believably simulate the spread of political ideologies and biases through a virtual population and how to present the effects of this simulation in a legible way to human users. We present *Lyra*, a simulation of a virtual town of characters that have varying degrees of political affiliations and ideologies modelled on the US political system. For virtual characters to be socially adept and add to the experience of the player, they must have a sizable expressive range of conversational repertoire. We advance an abstract knowledge base for the characters that groups various objects of discussion under overarching topics, tracks the sources from which they originate with their inherent bias or ratings, and allow the non-player characters (NPCs) to form opinions based on individual preferences or cultural norms. Our system can track the spread of influence (adverse or otherwise) and change in the views of the participant NPCs. Through a series of interactions with one another, the characters engage in conversations about current news articles on the topics of gun control and immigration. Characters attempt to sway one another towards their dispositions, they learn what topics of discussion are considered sensitive, or could add to growing antagonism or acceptance for themselves and their views among their fellow conversationalists.

We demonstrate our system with a case study that showcases a series of conversations where virtual characters discuss current political news from the U.S., exchange their views on individual news articles or issues of interest, and reevaluate their political ideologies and affiliations over time. For instance, an NPC growing up in a more liberal society may eschew conservative ideals, and have a low opinion of the same. Our simulated NPCs are aware of the difference in their internal attitude on a topic of discourse as well as the public opinion shared by other NPCs during their interactions. These differences can lead to the NPCs changing their attitudes over time or expressing opinions different from their attitudes to conform to the society they reside in over time.

We evaluate the believability of the simulation's depiction of the change in the characters' opinions with a human-subjects study deployed online. Our study has two sections, the first summative, evaluating the conversations and the virtual conversationalists themselves; the second formative, evaluating how such conflicts in opinions could affect future relationships and interactions the characters conduct. We evaluated the simulated conversations on a Likert scale ranging from 1-Not Believable at all to 5-Very Believable. We discovered the discussions had a mean believability rating of 3.3. Additionally, the human participants in the study were found to ascribe humanity to the actions of the virtual characters, describing agents that seemed to them to be "competitive" or that felt "marginalized", or discussing how "persuasive" characters seemed to be. We believe that these results support our hypothesis that Lyra can produce believable social conversation simulations (Togelius et al. 2013) with good explanatory results of the social mechanisms at play.

We imagine that in the future our opinion model could be used to evaluate how a virtual society would integrate and accept new additions with new members learning of the views and opinions of the society while bringing with them new ideas and concepts from their own culture. Similarly, opinion modeling for virtual characters could be used to study the spread of debatable ethical or moral influence and media bias. Characters could choose to accede to peer pressure (from the media or society) and change their behaviors in order to feel a mix of both private acceptance (that they are acting based on their views) and public conformity (to gain acceptance by the group). We believe the behaviors resulting from virtual characters modeled by this system would be more believable and improve a player's interactive experience.

Our work represents a step towards a better understanding of the mechanisms behind social influence and opinion dynamics, enabling more robust social intelligence and more believable social simulations. In summary, this work (1) overviews the previously established Lyra system (2) describes the design process and generation of conversational metadata (3) evaluates the generated conversations with a human subject study for their believability (4) extracts insights from the study to inform future research on how contentious discussions could affect social relationships amongst NPCs to more believably simulate the spread of opinions.

3.2 Related Work

In this section, we first describe related work from the narrative domain on believable virtual characters. Next, we discuss group formation from the perspective of social scientists, and psychologists to understand how believable virtual characters could be modelled to respond to group (or societal) archetypes and opinions.

3.2.1 Believable Non-Player Characters (NPCs)

There is no generally agreed-upon definition of believability. Instead within the narrative field, believability is used linguistically to describe that which is believable by someone. In terms of virtual characters, this could imply some aspect of their viewed interactions (either with the player or with each other) is believable. Togelius et al. (2013) describe how games that incorporate believable elements can elicit particular emotional responses to a player. They discuss how the generation of believable, human-like opponents lead to increased player enjoyment. Additionally, rich social interactions among NPCs have been found to improve the believability of interactive narratives and the player experience (Afonso and Prada 2008; Swartout et al. 2006).

With the wide-scale availability of mobile devices, and more recently the adoption of augmented reality (AR) technologies, researchers have manually authored narratives to document cultural heritage and community-based narratives or goals (Speiginer et al. 2015) as well as procedurally-generated narratives for various geo-locations populated with NPCs (Macvean et al. 2011; Dow et al. 2006; Leino et al. 2008). We posit that NPCs in real-world locations must be able to learn cultural, and societal values of the location they populate. Leeper and Slothuus (2014) build on prior work by Kunda (1990) discuss reasoning under partisanship (or motivated reasoning) stating a world devoid of partisan conflict is a dystopia. They argue that the novel contribution of motivated reasoning is the idea that individuals vary in the extent to which making accurate decisions is satisfying versus the extent to which they choose to reinforce their prior biases, attitudes or beliefs. Many traditional narrative planning systems allow for the former, with virtual characters able to create robust plans to achieve their goals (Cavazza et al. 2002; Young 2000). Towards this goal, our simulation allows for an NPC to evaluate their convictions over time, attempting to reconcile the disparities in their attitudes and beliefs with those of the other NPCs they interact with via conversation.

A key challenge posed by characters in a game is their ability to reflect their goals, personalities, and beliefs through dialogue or expositions. Rowe et al. (2008) describe how a requirement of the dialogue from a character must be that it is appropriate for the character personalities and preference while taking into account the narrative context and history. With this work, we do not directly address the natural language content generation of the conversation. Our system instead produces modifiers and keywords that state the intention of the characters and could be used to produce natural language dialogue utterances.

3.2.2 Group Formation

Group formation has been studied in depth by social scientists, historians, and psychologists to understand how humans respond to group (or societal) archetypes and opinions. When modeling group conversations, the physical or virtual space where conversationalists congregate can be used to contextualize the interaction, allowing us to incorporate the history, physical affordances, or cultural significance of the geographic location or the topic in question. Merely reading the news enables one to gain a perspective of humans forming groups to support various issues. These could be geographic groups, with articles describing how the *Scottish* voted to “overwhelmingly remain” in the Brexit vote; or political ideology groups, with reports on *Democrats* discussing immigration resolutions; groups based on shared interests, with news on *Whovians* that approve or condone representation of women in Doctor Who (Jowett 2014); or by grouping an occupation, with articles describing how *Tech executives* are contrite about election meddling. Latour discusses how individuals relating to one group or another is an ongoing process made up of uncertain, fragile, controversial and ever-shifting ties (Latour 2005).

3.2.3 Bias

Bias, in algorithms or decision making, can impact government, businesses and personal lives. Studying the impact of bias has recently become an emerging trend to study in computer science (Budak et al. 2016; Entman 2007; IBM Research 2018). Entman describes how the term can apply to news that distorts or falsifies reality (distortion bias), or news that favours one side rather than providing equivalent treatment to both sides in a political argument (content bias), or even with respect to the motivations and judgment behind decision-making processes (decision-making bias) (Entman 2007). Entman describes how studying media bias can provide insight into how the media influences the distribution of power. However, the bias in the media is not necessarily all bad. AllSides notes on those media outlets ranked with Center biases may leave out valid arguments from the left or right perspectives (AllSides 2018).

In Social Role Awareness (Prendinger and Ishizuka 2001), agents choose conversational responses based on their perception of their roles within the social context. A secretary addressing her manager could be more polite and responsive than one addressing an aspirant visiting the office. PyschSim (Pynadath and Marsella 2005) models influence amongst group members by examining how participants in a conversation view their relationships with one another and their beliefs and motivations about the world. Other work has virtual characters sharing their knowledge or gossiping about the world with one another with their bias (Evans and Short 2014). The most significant differences between our approach and the works previously mentioned are that our agents can reevaluate their biases or changes over time by subscribing to new opinion pools from their peers or other sources of information. We hope our model allows for a more natural conversation flow, with agents advancing and modifying their opinions over time. We hope our system will add to the believability and behaviors generated by these works by providing further motivation for character relationships and interactions.

Our work aids these efforts by endeavouring to use existing computational social psychology models to simulate how humans respond to and make decisions when faced with authority bias, or even how they respond to the cheerleader effect in conjunction with our social simulation.

3.2.4 Social Simulation

We argue that our research is a step towards machine enculturation (Riedl 2016) by simulating a society of virtual characters that have a predisposition towards learning new knowledge, cultures, and values based on their past interactions with both family (nature) and other societal influences (nurture).

Extensive research has been conducted on social rules and interactions between virtual characters. Versu (Evans and Short 2014) shows characters interacting with one another using pre-constructed social practices templates. These templates are constructed manually, can be time-intensive and require domain knowledge. Similarly with CiF, in Prom Week (McCoy et al. 2011a) the authors describe a social physics architecture model that constrains how NPCs behave. With their Actor-Network Theory (ANT) Latour discusses how individuals relating to one group or another is an ongoing process made up of uncertain, fragile, controversial and ever-shifting ties (Latour 2005).

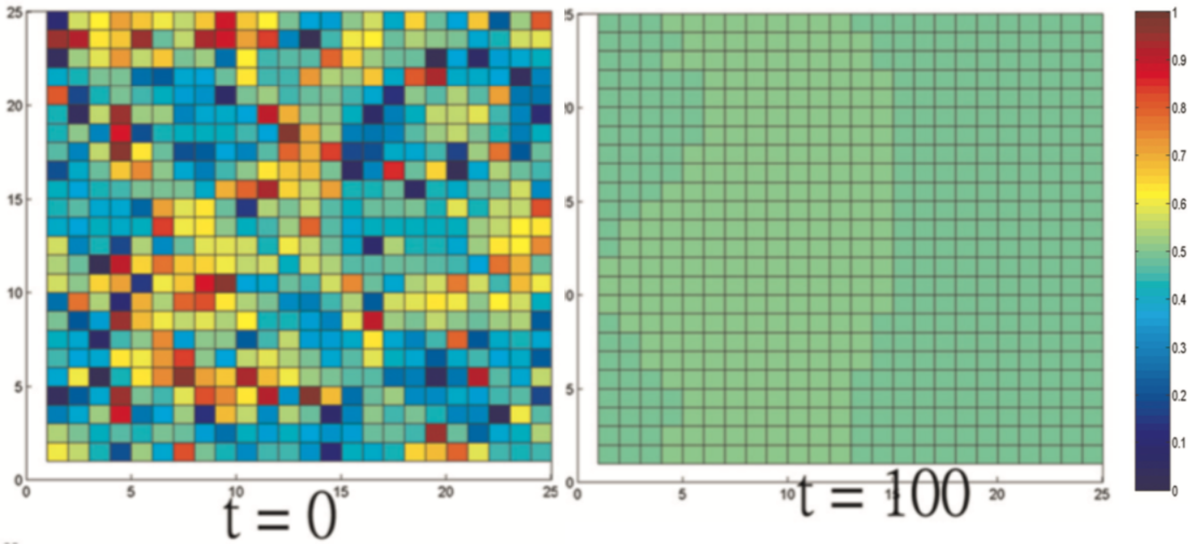


Figure 3.1: An evolution of agent attitude dynamics represented by cellular automata (Wang et al. 2014). The left graph shows initial variation in opinion and the right graph shows the more homogeneous opinions after 100 iterations.

Our simulation consolidates these two approaches, that of ANT and the traditional narrative intelligence approach. Virtual characters' group membership changes over time based on their recognition of their internal attitudes and the opinions of the society around them. With this approach, rather than manually authoring social rules and beliefs (as in systems like Versu (Evans and Short 2014)), social rules emerge organically over time as beliefs and attitudes that go against the group's values would be looked upon unfavourably by its members. We believe this would reduce the authoring burden of the social rules or templates (Evans and Short 2014; McCoy et al. 2011b), allowing for interesting emergent gameplay.

Lyra extends current theories of dynamic opinion modelling research (Wang et al. 2014; Asch 1955) with the goal of being able to model societies with NPCs capable of exploring complex issues of politics, religion, making decisions, and forming social relationships based on their views. With Wang et al. (2014), agents are modeled as individual nodes in a social network represented as a cellular automata. Agents may exchange opinions with other agents surrounding them. Wang et al. defines how agents every agent's feelings on a topic is informed by an inner **"attitude"** towards the topic that cannot be perceived by other agents, an outward expressed **"opinion"** and the level of **"uncertainty"** they feel about their opinion. Agents may adjust their internal attitudes or express modified opinions from their attitudes, on hearing the opinion of other agents in the surrounding cells (Wang et al. 2014).

Due to space restrictions, we refer readers to the Wang et al. (2014) paper, and the Asch (1955) paper for the details on their experiment. We recognize that the threshold values and model evaluated in the Wang et al. paper may not exactly conform to an exhaustive list of objects of discussion or topics of discourse. However, their proposed agent model combines normative social influences with a continuous dynamics model in a novel approach. Our objective is to extend these current theories of dynamic opinion modeling research to the narrative intelligence community with the goal to simulate virtual societies capable of exploring complex issues of politics, religion, or even simply movie ratings. Towards this goal, our contribution builds on that of Wang et al.'s in the following ways:

- Prior work fails to model the complex and ever-changing social relationships between conversationalists. The authors assume a grid-based society where the same neighboring agents surround an individual throughout their simulation. Our method proposes a more utilitarian definition of social relationships where NPCs with differing or similar opinions could change relationships over time, allowing their old social connections to dissolve over time.

- Instead of a single object of discussion, we allow characters to discuss a variety of information clustered by topics. This allows for relationships where characters that agree over a few views but disagree over others to change their affinity for one another over time.

- We allow for the simulation to add new concepts and topics over time. We believe this could lead to virtual characters to extend their knowledge base while retaining their individual views on existing knowledge.

3.3 Goals

My goals for Lyra were divided into two sets. System goals, that helped define the architecture and requirements of the Lyra system; and Evaluation Goals, that constrained the output of the system, and allowed us to extract insights and evaluate Lyra conversations for believability.

3.3.1 System Goals

My system goals for Lyra defined the structure of the Lyra Project (and the order of the following sections), as well as the design constraints and requirements for the system. We list our goals for the Lyra simulation design as follows:

- **G1:** Account for bias in characters where agents may have a predisposition to adopt a specific view from prior experience.
- **G2:** Account for bias in the information. Information and sources producing information may have an inherently biased perspective.
- **G3:** Ability for characters with similar opinions to form relationships, and allow ad-hoc groups developing during social interactions to discuss their opinions on various topics.
- **G4:** Be able to use the same discussion model for a variety of different data sources to simulate opinion modeling on discussions.

In Section 3.4 I describe the Lyra system, and established our model of world knowledge that can take into account biases associated with the knowledge and its source. In Section 3.5, I discussed the model of the characters internal attitude and expressed opinions on the topic based on prior work that models self-perception agents. Finally, in Section 3.6, our simulated algorithm for simulating discussions, to form ad-hoc groups to discuss their views and closer relationships with characters that had similar perspectives is discussed.

3.3.2 Evaluation Goals

With our evaluation goals, we expand on our earlier goals by adopting the following goals regarding the evaluation of the Lyra System.

- **G5:** To generate descriptions of the change in opinions of the conversationalist NPCs that allow readers to follow an NPC's reasoning.
- **G6:** To evaluate these generated conversations with a human subject study for their believability.

- **G7:** To extract insights from the study that can inform future research on how contentious discussions with polarizing views could impact NPC social intelligence, and more believably simulate the spread of opinions.

These evaluation goals described the remaining structure of this project. We describe steps to achieve G5 in Section, *Designing Legible Simulation Output*. Likewise, the study design and approach for G6 can be found in the *Study Design* section. Finally, for G7 we described the results from our study in the *Analysis* and *Discussion* sections where we analyze study results to answer four research questions that can help guide future research on character believability.

3.4 Lyra's Model of Knowledge

Our knowledge model describes how information in the simulation world is structured. This can be overviewed as follows. For a single discussion, the participants in the discussion choose an *Object of Discussion* to converse on, obtained from a *Source*. The Source and the Object of Discussion are associated with a *Rating*. Multiple objects of discussion can be clustered to form a *Topic*.

Topics	**Objects of Discussion**	**Source**	**Rating**
Political Issues e.g. Immigration, Gun Control	Individual news articles	Online or Print Media	Political Bias or Affiliation
Political Issues e.g. Immigration, Gun Control	Political candidates	Interviews, Candidate Rally	Approval Rating
Research Topics e.g. AI, Games	Conference Papers	Journals, Conference Proceedings	Journal or Conference Rankings
Film Genres e.g. Horror, Sci-Fi	Movies	Movie Studios	Rotten Tomatoes score

Table 3.1: Examples showing how discussions can be simulated on various datasets using the proposed knowledge model

Our model of the knowledge base can be used for a large variety of datasets while affording the same discussion and opinion modelling. For instance, simulating debates among NPCs about *current news articles* clustered by *political issues* and ranked by their *bias*. Similarly, we could use our model to discuss the merits of various *journal articles* clustered together by *research topics* and ranked by *journal rankings* or have audience members discuss their *movie preferences* clustered by *movie genres* and ranked by their *Rotten Tomatoes rankings*. Some datasets considered during the design phase of this model have been highlighted in Table 3.1.

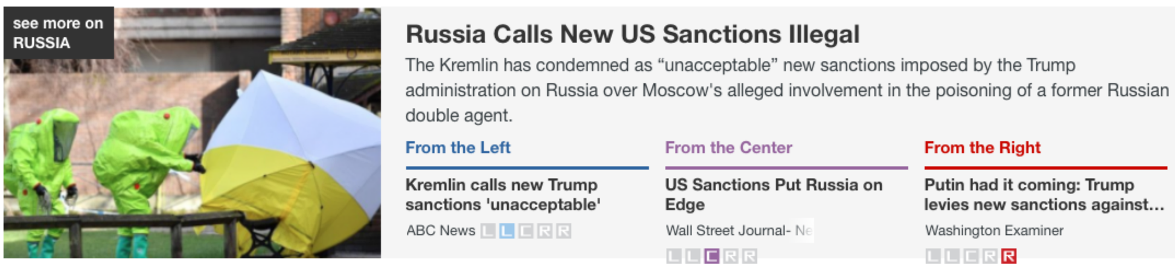


Figure 3.2: Case Study: Example of a topic of discourse, Russia, and some news articles associated with it, each labeled with their own media bias (AllSides 2018)

Our simulation uses a corpus of news articles from AllSides.com (AllSides 2018) that use a combination of blind bias surveys, editorial reviews, third-party research, independent research, and community votes to calculate media bias of the information. An example of a topic of discourse and the news articles surrounding the same has pulled from the site and shown in Fig. 3.2. We see three articles, each labeled with it's source and it's bias.

3.4.1 Ratings

Ratings are defined as the value of the information learned by the NPC in the system. This rating could represent either (1) the personal judgment or favour associated with the presentation of the information, or (2) a measure of the impartiality of the unit of information.

3.4.2 Topics

Topics are a clustering of information regarding a specific subject, or field of information. A specific information unit can be a part of multiple topics at the same time. For instance, a discussion of procedural content generation could belong to the topics of both artificial intelligence or game design.

3.4.3 Objects of Discussion

This single unit of information forms the basis of our discussion model. While interacting with one another, virtual characters search through their knowledge base and conversational repertoire, choosing a single object of discussion to debate. An NPC that adds a new object of discussion to his knowledge base will note the original authorial rating intended to be affiliated with the information, and associate with it their own opinions on the topic. These views could be based on prior discussions of the information with conversationalists that introduce the character to the information, as well as on the character's current view of the topic to which the information belongs.

3.4.4 Sources

Sources are creators of knowledge. They may create information covering a wide variety of objects of discussions and topics. Sources may also have associated with them a rating, representing the expected rating of the information they produce. NPCs may use this rating to choose to subscribe or unsubscribe to these over time based on their current inclinations.

3.5 Social Character's Views

Every participant in the discussion has their own *Bias* and *View* on the information and can express their opinions on the object of discussion at hand. These elements and our dataset have been described in further detail below. The attributes of an agent's view are modelled based on those by Wang et al. (Wang et al. 2014).

We represent these NPC views as consisting of an *Attitude*, an agent's private views on a specific object of discussion, an *Opinion*, an agent's outwardly expressed or shared views, and a *Uncertainty* about their views. Additionally, we use two thresholds, a *Public Compliance Threshold* which describes when the agent chooses to comply with the public opinion to feel accepted within the community, and a *Private Acceptance Threshold* which describes when an agent will choose to stand by their views. Finally, we define a *Bias* to be the agent's predisposition to adopt a particular leaning (left/right) on a topic in a discussion.

3.5.1 Bias

Bias is the agent's predisposition to adopt a particular view on a topic in a discussion. This bias is informed by either (1) the agent's views inherited from their parents or (2) a mean of their views on all objects of discussion under the said topic or (3) the initial bias they learn from the conversationalists when the topic was added to their knowledge base during a discussion.

3.5.2 Attitude (att)

Attitude is the agent's private views on a specific issue. Attitude is a real number in the range $[-1, 1]$ and represents an evaluation of the object of discussion.

3.5.3 Opinion (op)

Opinion is an agent's outwardly expressed or shared views on a specific issue. Like attitude, opinion is a real number in the range $[-1, 1]$ and reveals the agent's opinion on the object of discussion to the other dialogists. There may be a discrepancy in the attitudes and opinions of the character since a character may not represent their attitudes accurately to participants. A human example of the situation where this is apparent can be seen in examples of an employee in conversation with his managers who choose not to express his disagreement to avoid being punished.

3.5.4 Uncertainty (unc)

Uncertainty is a measure of an agent's confidence in their view. The higher the uncertainty, the more likely the agent is to change his mind or be accepting of other perspectives. As an example, an NPC may express opinions about the legality of abortion in their town. However, the agent may have lower confidence in their attitude if (1) information in their existing knowledge base inadequately back them, (2) if contradictory opinions are presented to the agent with high certainty, or (3) if the agent is surrounded by a society a majority of whom disagrees with him. unc is a real number in the range $[0, 1]$.

3.5.5 Public Compliance Threshold (pub_thr)

When the strength of the public opinion exceeds pub_thr, the agent will choose to comply with the public opinion to feel accepted within the community.

3.5.6 Private Acceptance Threshold (pri_thr)

If the strength of the public opinion, or pri_thr, is below this value, the agent will choose to stand by their views. The pri_thr is a real number in the range $[0, 1]$. Professors or experts on a particular topic in our simulation have higher values to indicate their expertise.

3.6 Simulation of Discussion

Our model accounts for the fact that the same participants could have different opinions (and therefore social relationships) based on their shared interests in other discussion topics, such as computer science, or hiking. This allows for relationships where characters that agree over a few views but disagree over others to change their affinity for one another throughout multiple discussions.

We begin by clustering similar expressed opinions of all participants of the conversation using the Jenks Natural Breaks Optimization method (Jenks 1967). This mirrors how humans interact. For instance, a group of fans may congregate at a water cooler at work, forming coalitions of people that argue about who should rule Westeros (Benioff and Weiss 2019). The number of opinion groups formed indicates whether a *public opinion* on the matter has developed and the presence of normative social influence (or peer pressure). The fewer the number of clusters that form, the more likely it is that an agent who maintains their views contrary to public opinion will feel rejected (Wang et al. 2014).

3.6.1 Public Opinion formed

We calculate each agent's change in views based on their certainty and the strength of others' views. Agent's with high uncertainty in their views are more likely to accept the public opinion and their views are modified accordingly. If the agent has low uncertainty, we find the largest clustered opinion group with views closest to that of the agent. We then calculate the public opinion strength for the selected group and decide if an agent's attitudes or opinions are affected. The strength of the public opinion as perceived by each agent is affected by:

- The size (f_a) of the group. The larger the group, the stronger the public opinion.

$$f_a = \begin{cases} 0,\text{if } x_a \leq 1 \\ x_a/10,\text{if } 1 < x_a \leq 10 \\ 1,\text{if } x_a > 10 \end{cases}$$

- The homogeneity (f_b) in the opinion of the group defining if the group come to a consensus

$$f_b = 1/(1+e^{24x_b-6})$$

- The discrepancies (f_c) in the agent's opinion and attitude.

$$f_c = 1/(1+e^{-12x_c+6})$$

Next, the agent measures their own uncertainty with the strength of the public opinion by calculating two threshold values, $th_1 = 1 - agent.unc$ and $th_2 = \max(0.6, th_1)$.

- Low Opinion Strength ($op\_str < th_1$): If the opinion strength is too weak, the conversationalist does not change their mind, recognizing the discrepancy between their internal attitudes and ideas and those of the group.
- Moderate Opinion Strength ($th_1 \leq op\_str < th_2$):
  - Members with a low uncertainty find the opinion strength of their group strong enough to modify their opinions to the mean of the group. Agents then find their internal attitudes, and their expressed behaviours are inconsistent, and so change their attitudes to match. In this case, agents believe that the change in their views is a natural and expected evolution, and do not realize they are bending to public opinion.
  - Agents with large uncertainty realize that they are conceding the discussion, and bending to public opinion. They change their external opinions and internal attitudes to match.
- High Opinion Strength ($op\_str \geq th_2$): The agent realizes the strength of the opinion. In this case, the agent may choose to conform to the public opinion with their outwardly expressed views and change their opinion to the mean of the group.

However, they *do not* change their inner attitudes, and in the absence of external pressure will revert to their attitudes.

3.6.2 No Public Opinion formed

The agent finds the cluster of opinions with the opinions most similar to theirs. The NPC modifies their opinion to the mean of the cluster and their internal attitudes on the information being discussed.

Due to space restrictions, we refer readers to our prior work (Azad and Martens 2018) for further details of the algorithm and simulation described above.

3.7 Case Study: Political Ideologies

In this divisive age, it is difficult (yet unavoidable) to discuss current political events with family or friends. APIs for major media sources are available with access to news articles on various topics. As a case study, our simulation uses a corpus of news articles (AllSides 2018), grouped by their political issues. Characters are initially assigned political affiliations and biases. The rating system, in this case study, is based on that of the U.S political-ideological system. For the simulation, in the beginning, characters are subscribed to sources that confirm their political bias. For instance, a Centrist NPC may subscribe to the *Associated Press* as a news source.

News Source	**AllSides Media Bias Ranking**
New York Daily News	Left
New York Times	Lean Left
Associated Press	Center
Boston Herald	Lean Right
Fox News Editorial	Right

Table 3.2: Examples of the AllSides Media Bias Rankings obtained for NPC subscriptions to media sources

3.7.1 Our Dataset

While our system can be used to simulate a variety of different conversational topics (as shown in Table 3.1) we needed to choose a domain that would allow for a post-simulation human subject evaluation.

Problem: Choice of Conversational Domain

The Lyra knowledge model can be used to simulate conversations in a variety of domains while affording the same discussion and opinion modelling (see Table **??**). With this study, we needed to choose a familiar domain where our target demographics could imagine accompanying dialogues and be able to relate to the forming of clusters and coalitions of like-minded NPCs. Additionally, respondents should be able to judge the NPCs in swaying others to their perspectives for their believability.

Solution: Political Domain Chosen

Our reasons for selecting the US Political System as our chosen domain were threefold. Firstly, this subject matter was considered to be familiar and relatable for our target survey demographics. Next, the range of political stances on the topic have familiar, quantifiable metric (see Fig. 3.6). Finally, the topic could elicit inferences of plausible dialogue occurring amongst characters based on the respondent's own experiences of past politically charged conversations. This would enable respondents to better judge our generated conversations for believability. For this study, we limited the topics of discussion in the domain to *Immigration*, and *Gun Control and Gun Rights*.

Rating

We use media bias as our rating and associate with each bias a value as follows: *Left*(-1.0), *Lean Left*(-0.5), *Center*(0.0), *Lean Right*(0.5), *Right*(1.0). The bias ratings in our dataset are obtained from AllSides using a combination of blind bias surveys, editorial reviews, third-party research, independent research, and community votes to calculate media bias of the information (AllSides 2018) as can be seen in Table. 3.2.

Topics

We use U.S. Political Issues such as Civil Rights, Immigration, Healthcare, Free Speech, Gun Control, and Abortion (AllSides 2018) each with an equal number of articles representing every bias.

Objects of Discussions

Individual news articles are our objects of information. A character will note the original authorial bias of the information and associate with it their views based on their current attitude towards to the topic, their overall political affiliations, and their discussions on the article with other conversationalists.

Sources

Sources are media sources that publish articles on a wide variety of issues. NPCs may subscribe or unsubscribe to these over time based on their current political inclinations.

Overall Political Affiliation

is a weighted average of the agent's attitudes of all topics in the agent's knowledge base (ranked by an agent's priorities). For instance, a simple measure how Liberal or Conservative a person is could be expressed as a weighted average of their attitudes on the topics of gun control, abortion, homosexuality, tax reform, and so on.

3.7.2 Social Interactions and Discussions

We simulate a town where characters can interact with one another. Our preliminary experiment allows for two types of organizations, Schools, and Businesses, to facilitate group discussion.

Schools

Schools choose a subset of topics from the world to teach their students. Professors are modeled to have a low uncertainty value regarding their views. This in combination with the fact that they are regarded as authority figures in the simulation implies that a student is more likely to adopt their views. In Fig. 3.3 one can see the knowledge base of a recent graduate after he reevaluates his views on Immigration.

```
Character: Horacio Plyler
Evaluation of views on Topic: Immigration

Knowledge Base:
    Topics: [Immigration, EPA]
    Number of Sources: 27
    Number of Articles: 43

Old View --> att: 0.884 | op: 1.0   | unc: 0.1156
New View --> att: 0.010 | op: -0.33 | unc: 0.3401
```

Figure 3.3: The political news and opinions knowledge base for a character that graduated from school

Businesses

NPCs may apply to work at open positions in various local businesses. The application to these positions is based on the knowledge as well as the opinions an NPC acquires over time. For instance, an NPC may be required to have specific views on the topic of abortion as a qualification to work at a local hospital that matches those of their colleagues.

```
Discussion for 11 minutes on "Room for Debate: Should `Birthright Citizenship' Be Abolished?"
Source: NYTimes
Participants: Richard Cain, Ruth Franklin, Vickie Nguyen-Self, Suzanne Sorenson

Views:
Richard: less uncertain.                            Ruth: no change in views.
Old --> att: -0.5 | op: -0.559 | unc: 0.632         att: 0.0  | op: -0.063 | unc: 0.225
New --> att: -0.5 | op: -0.559 | unc: 0.554         att: 0.0  | op: -0.063 | unc: 0.225

Vickie: change in views, less uncertain.            Suzanne: no change in views
Old --> att: 1    | op: 0.535  | unc: 0.464         att: -0.5 | op: -0.559 | unc: 0.059
New --> att: 0.948| op: 0.521  | unc: 0.497         att: -0.5 | op: -0.559 | unc: 0.059
```

Figure 3.4: Discussion generation involving four agents discussing an NYTimes article.

Sample Discussion Outcomes

We decode in prose a typical outcome for a discussion from our simulation as shown in Fig. 3.4. NPCs discuss an article titled *"Room for Debate: Should 'Birthright Citizenship' Be Abolished"* at work with colleagues. The article falls under the topic of *Immigration* and is published by the source *NY Times* with an original authorial bias calculated by AllSides as *Leaning Left*. The duration of the discussion is *11 minutes*, representing the number of times the algorithm is run, and the views of the participants are updated.

Ruth and Suzanne learn about the article for the first time. They choose to accept the outcome of the discussion as their opinion after applying any pre-existing bias on the topic of Immigration. Richard, whose political views Leaned Left ($att = -0.5$) before the discussion, is more convinced about his views after reading the article. As such his uncertainty on the subject reduces, but his views stay the same. Vickie, whose political views were aligned Right ($att = 1.0$) before the discussion changes her views slightly over the course of discussion ($att = 0.948$) and finds herself a little more uncertain about her view on the article.

However, since the internal attitudes of all four participants on the article and the topic of Immigration (not shown in Fig. 3.4) remain the same, their overall Political Affiliations do not change… yet.

3.8 Designing Legible Output

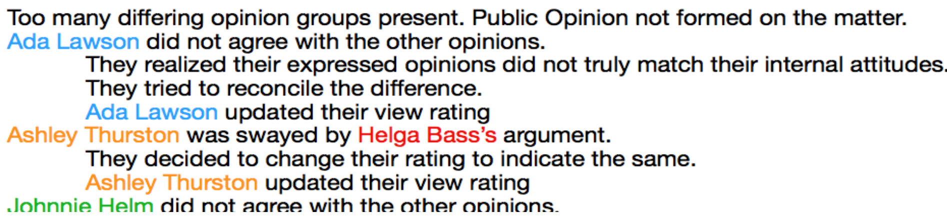
A critique of our earlier system lay in readers having difficulty understanding and producing explanatory descriptions of how and why characters changed their mind over time. A sample output from our earlier system can be seen in Fig. 3.4. We redesigned the simulation output to be presented to the reader in discrete rounds. In this section, we describe our design process for creating legible simulation output to human readers.

3.8.1 Problem: Authoring Bias for Dialogues

Authoring accompanying dialogue to match the views of the characters per conversation round was found to be untenable. It was not our intention to author the natural language content of the opinions proffered by the characters during the rounds. Given the thesis of this paper, any human authoring of content would need to be rated for the bias of its author and the content.

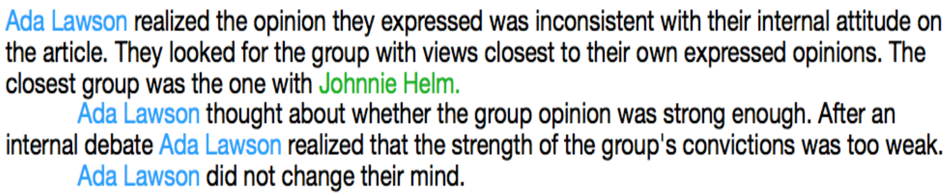
3.8.2 Solution: Designing Textual Descriptions

To circumvent the authoring bias problem, we generated descriptions of these conversation choices that would allow the virtual characters to explain their internal state, actions taken, and any changes in their attitude without the content of the opinions being shared. We a



Too many differing opinion groups present. Public Opinion not formed on the matter.
Ada Lawson did not agree with the other opinions.
They realized their expressed opinions did not truly match their internal attitudes.
They tried to reconcile the difference.
Ada Lawson updated their view rating
Ashley Thurston was swayed by Helga Bass's argument.
They decided to change their rating to indicate the same.
Ashley Thurston updated their view rating
Johnnie Helm did not agree with the other opinions.

(a) Round 1 of discussions



Ada Lawson realized the opinion they expressed was inconsistent with their internal attitude on the article. They looked for the group with views closest to their own expressed opinions. The closest group was the one with Johnnie Helm.
Ada Lawson thought about whether the group opinion was strong enough. After an internal debate Ada Lawson realized that the strength of the group's convictions was too weak.
Ada Lawson did not change their mind.

(b) Round 2 of discussions

Figure 3.5: Excerpts from a generated conversation

sample conversation excerpt in Fig. 3.5 depicting a round of a conversation among 4 NPCs at a school. In the excerpt, Ada realized they were experiencing cognitive dissonance, and chose to reconcile the perceived difference between their internal attitude and the opinion they expressed to other characters.

3.8.3 Problem: Following the Change in Character Views

A critique of the earlier version of Lyra was that it was hard to follow the change in a character's views over time. While the final political affiliations and opinions can be seen in Fig. 3.4, it was hard for readers to understand what a conversation between these characters could look like, or evaluate whether these changes were believable.

3.8.4 Solution: Our Simplified Political Rating Scale

To make the change in the character's opinions more visual, and easy to relate to we used a simplified rating system for the political affiliation of the virtual participants. All Graphs summarizing the conversation for the participants used this scale going from -1, representing "left" on the political spectrum, to 1, representing "right" on the political spectrum.

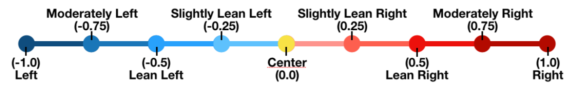


Figure 3.6: Simplified political scale for each topic discussed

Further, we described both the Media Bias and the Character Attitudes on topics on the spectrum using descriptions for the positions from Allsides.com (AllSides 2018). These descriptions, provided to the survey respondents, have been added to the Appendix at the end.

3.8.5 Problem: Lengthy Textual Descriptions

Initial practice runs of the survey made it apparent that our subjects found it difficult to track all the variables mentioned (for instance, attitude, opinion, uncertainty, familiarity with the topic, etc) described in the conversation text.

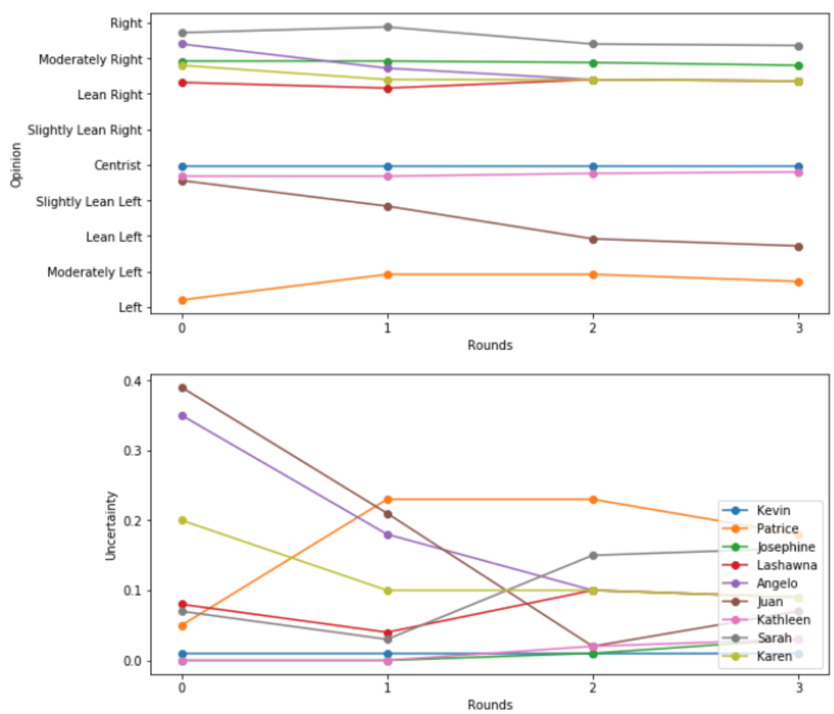


Figure 3.7: Summary of the change in the opinions of the characters over 3 rounds of discussion.

Solution: Graphical Descriptions

We supplemented our textual descriptions of the conversation with two summary graphs that showed the swing in the opinions and the swing in the uncertainty for the characters throughout the conversation rounds (see Fig. 3.7).

3.9 Study Design

To understand Lyra's effectiveness at believably simulating opinion propagation and the social dynamics of politically charged conversations, we conducted a human subjects study asking readers to read simulation output and answer questions in a survey. In this section, we describe our survey procedures and analysis process.

3.9.1 Procedures

Our survey asked questions to determine participants' political affiliations and biases, the news media sources they subscribed to, and how differing opinions affected their social relationships. Next, they read 4 computers generated conversations between groups of virtual characters with different political ideologies and biases (see Fig. 3.5) and looked at charts summarizing the rounds (see Fig. 3.7). They were then asked questions regarding the believability of the conversations, and the intentions of the virtual characters participating in said conversations. They were also asked to rank the persuasiveness, and feelings of membership or inclusiveness with the group for each character. Participants were given the option to enter open text for the conversations for additional feedback or take-aways. Finally, they were asked to fill out a short demographic form.

The survey was distributed online via email lists and social media. The first 25 participants that completed the survey were offered to be paid with an Amazon Gift card. The survey took about one hour to complete.

3.9.2 Response Demographics

Our survey had a total of 21 respondents. Of the respondents, 11 identified as male, 8 identified as female, 1 participant chose to describe their gender differently, and 1 declined to respond. When asked about their education, 11 had completed their Master's degree, 4 had completed their Doctoral Degree, and 4 had completed their Bachelor's degree, a participant had an associate degree and another had some college credit but no degree. Of the surveyed, 17 were between the ages of 25-34, and 4 were above the age of 35. 16 of the 21 participants identified with the *Liberal* political descriptor, 4 identified as *Conservative*, and one declined to state a political affiliation.

Coder	Code	Description	Sample matched survey quotes
1	Influence by a Group	Identification of a group influence on NPC opinions	She was swayed by the rest of the group
1	Influence by personal bias of participant	The comment seems to have been influenced by the participant's own personal bias	The centrists didn't change at all; which doesn't seem characteristic of the topic
2	Decreasing Certainty	Observation of characters' decrease in certainty	The fluctuation from high certainty back to uncertainty in a seemingly short time period.
2	Liberal Open Minded	Stating the belief that liberal people are more open minded or right-wingers are less likely to change their minds	That the most liberal person would be the person most open to changing their mind

Table 3.3: A few codes from the Initial Coding Scheme

3.10 Method: Qualitative Analysis

With this section, we detail our method for the qualitative analysis of the survey results.

Constructing Queries

We chose to use a directed approach to content analysis in both phrasing and analysis of open text queries asked during the survey (Mayring 2004). Our goal was to be able to validate and extend conceptually our theoretical framework and model for opinion dynamics amongst NPCs. Thus, our queries were framed to probe participant predictions and expectations of the conversation and explore their understanding of the relationships between our variables of interest. Primarily these included the believability of the conversation, the change in the opinions and attitudes of the participant NPCs, the relationship between uncertainty and change in opinions. Keeping this in mind, we asked participants for open text responses to four questions as detailed below:

- What was the most believable part about the conversation described above?
- What was the least believable part about the conversation described above?
- One reasoning question about an NPC per conversation:
  - Why do you think Ashley was so uncertain of their views?
  - Why do you think Ada's uncertainty reduced?
  - Why do you think James's uncertainty increased?
  - What does Juan's change in opinion tell you of their private attitude of the conversation?
  - Why do you think Amy's uncertainty increased after Round 2?
- Any thoughts of take-aways from this conversation that you would like to share

With the reasoning question, we hoped to incite responses to indicate to us the mental model of the respondent, and whether their interpretation and expectations of the change in an NPC's views matched those of our algorithm.

Directed Content Analysis

After the survey ended, the open text responses to the question described above were transferred to a Google Sheets document.

The authors of this paper performed initial open coding and used content analysis to analyze the data. Both coders were familiar with the underlying theory of the discussion model and the formulated research questions. This allowed the initial codes noted to have a more structured, directed approach (Hickey and Kipping 1996) described by Mayring (Mayring 2004) as deductive category development and application. The step model for this analysis has been depicted in Fig. 4.1 as shown below.

The data was read from start to end to obtain a sense of the whole. First impressions, and thoughts were noted down to capture key concepts based on the variables of interest (Potter and Levine-Donnerstein 1999). Both authors were aware of the variables of interest for the study, namely, to better understand the uncertainty, attitude, opinions and believability

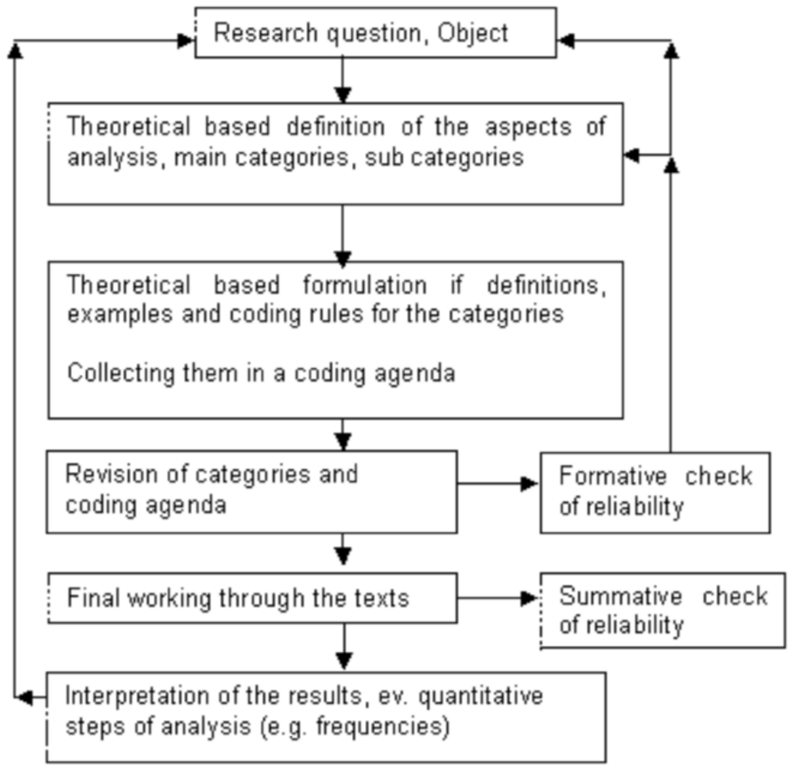


Figure 3.8: Step model of deductive category application (Mayring 2004)

of the conversations. This allowed the authors to create two independent codebooks with explicit definitions, examples and coding rules for each deductive category discovered. A few codes from this initial coding scheme (or codebook) have been described below in Table 3.3.

The authors performed open-coding analysis independently, reading the responses to derive codes (Miles et al. 1994; Morgan 1993). An initial discussion was conducted to discuss, negotiate and merge the codebooks as a formative check for reliability. For instance, the phenomenon coded as "Influence by personal bias of participant" by Coder 1 and "Liberal Open Minded" by Coder 2 were found to be linked and were merged and renamed to "Used Political Stereotype" for the initial coding scheme formed. Additionally, during the discussion, other codes were organized into more meaningful clusters. For instance, the label "Increasing Certainty Expected" was decomposed into two sets of codes, the first, "Increasing Certainty", and "Decreasing Certainty", the second, "Expected" and "Unexpected". This allowed for the second set to be used in conjunction with other tags, to capture the survey respondents expectations, and allow for more in-depth frequency

Tag	Definition
#NPCMentioned Unprompted	The behavior of a character was noted when they were not mentioned in the question
#ChangedOpinion	Noting that an NPC or a group of NPCs had a change in their opinions
#StandingGround	No change in opinion. The NPC stood their ground.
#SimilarViewsConverge	Noting that NPCs with similar views eventually converge their views
#GroupInfluence	Noting when an NPC is swayed by a group.
#UsedPoliticalAffiliation Stereotype	Respondent made a stereotypical judgement about a political affiliation.
#IndividualInfluence	Noting when a character is swayed by other individuals (but the individuals are not identified as a group)
#DecreasingCertainty	Noting that an NPC's certainty in views decreased / uncertainty increased.
#InferFactsFrom	Infer facts (or make assumptions) that are not given to them by us
#CertaintyConvinces	Noting when a character who is more certain in their views has more influence

Table 3.4: High frequency codes and definitions obtained after qualitative analysis

analysis at a later point. Finally, this initial discussion allowed for the discovery of new codes that only one or the other coder had noticed without being biased. For instance, one author noted that several respondents discussed the existence of an "Overton Window", or that respondents noticed that NPCs with higher "Certainty Convinced Others", while the other author was interested in how respondents displayed an "Emotional Response" to the conversations they read, or that responses could be tagged to indicate whether the "Clustering" of NPCs during the discussion phase was found to be "Believable". This resulted in the creation of an initial coding scheme with 34 codes identified and described with their usage and examples.

Validating Initial Coding Scheme

To further establish rigour, reliability, reduce the coding scheme's discriminant capability – that is, reducing coding errors – and to validate this initial codebook, two additional independent coders were recruited. These new coders were unfamiliar with the project and were given a description of the research to explain the purpose of the same. Next, we discussed the Initial Coding Scheme developed, examples of the discovered codes, and what each one meant. Based on this discussion some definitions and examples were further clarified. Next, we selected 15% of the survey data at random from the Google Sheets. This was paired with the discussions and questions that the text was in response to and given to the new coders.

Measure	**Agreement Value**
Fleiss kappa	0.9099
Cohen kappa	0.9121
alpha	0.9012

Table 3.5: Interrater agreement

Each of the new coders coded the subset of data given to them using the codes in the Initial Coding Scheme developed. They were encouraged to create new codes if they felt that the existing scheme did not fully represent the data. One of the authors also re-coded the same segment using the Codebook provided. Next, we compared results and discussed problems where there were discrepancies in the codes used by the authors and the coders. Adjustments were made after negotiations, and some new codes were discovered, and old codes modified. For instance, the addition of "Opinion Changed Despite Certainty" to depict situations where survey respondents found that an NPC's opinion changed despite their certainty and "Meta Discussion" to capture discussions and feedback about the study design or survey were added. The final Thematic Coding included 44 codes. The Intercoder reliability was calculated using NLTK's Fleiss' Kappa and Krippendorff's alpha statistical measures to assess the reliability of agreement amongst the three coders (Fleiss 1971; Loper and Bird 2002). Additionally, Cohen's Kappa was calculated amongst the raters and the author (Cohen 1960) for the random 15% of the data analyzed. The results have been shown in Table 3.5. The coders were found to have *Almost Perfect Agreement*. Thus, the coding scheme was finalized, and used by the author to code the rest of the data before further frequency analysis was performed.

Thematic Codes

Some of the thematic codes that occurred with higher frequency in our analysis have been described in Table 3.4 above. The remaining codes with their descriptions and sample survey respondent quotes tagged with the thematic codes can be accessed in the appendix to this paper.

3.11 Results

In this section, we detail our research questions along with relevant insights produced by our analysis.

3.11.1 RQ1: Does the measure of the believability of the generated conversations depend on the personal political biases of the respondent?

We asked participants to rate their political bias on a left-to-right scale as well as to provide their results from the Pew Research Political Typology quiz (Pew Research 2017). Fig. 3.9 shows how liberal and conservative respondents rated the believability of the conversations. We found that the responses from Conservatives and Liberals were not significant ($p > 0.05$) in the believability rating of Discussion 4.

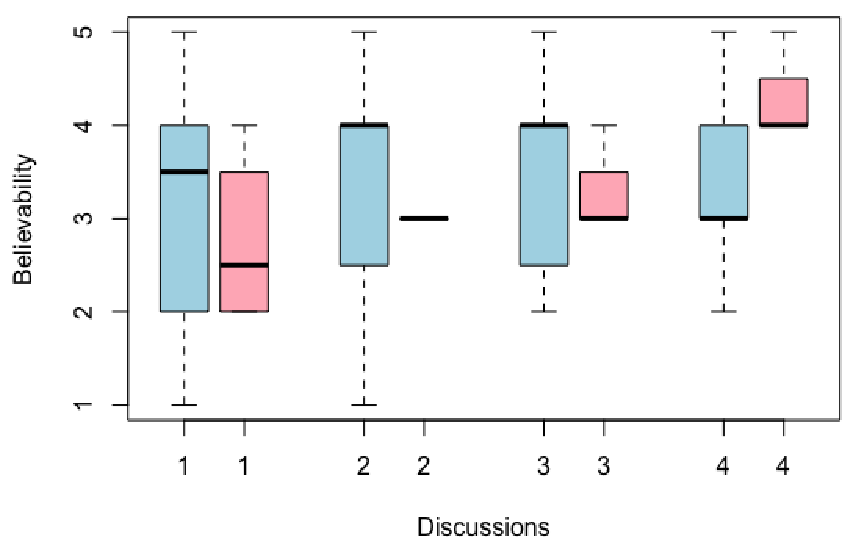


Figure 3.9: Perceived believability rating of the 4 generated discussions by Liberals and Conservatives (per Pew Research Political Typology results)

We hypothesized that the personal biases of the participants on the topics discussed by the NPCs would impact their believability ratings in the groups where those issues were discussed. To test this, we asked participants to *"Rate their views on a 5 point Likert Scale ranging from 1 (Strongly Left-wing) to 5 (Strongly Right-wing)"* on the topics of Gun Control, Legal Immigration and Illegal Immigration. These were topics discussed by the NPCs to see if their perspectives on a particular topic affected their suspension of disbelief in the generated discussions.

Since our data as not normally distributed, we used the non-parametric Mann-Whitney U test to compare the groups. However, the difference between the groups was not significant ($p > 0.05$). This implies the respondents' political preferences on a particular topic did not impact their rating. Interestingly, 3 of 21 participants' familiarity with the *topic* discussed influenced their experience and interpretation of the conversation. In a discussion generated with smaller variations in the views of the NPCs, one participant mentioned that *"[the fact that] people [would be] swayed by the other participants [wasn't] likely [to happen] with [discussions on] gun-control."* Another participant pointed out that the NPC, Juan's *"views on gun-control aligned with liberal views."*

Finally, we asked respondents to select all the political descriptors from a hand-generated list that they identified with. We ran a linear regression model and found that the political identifiers were not significant ($p > 0.05$). We found participants tended to project their own bias and experiences on to the agent while explaining why an agent made decisions, with statements such as *"Ada is a typical right-winger and is looking for viewpoints to confirm her own bias; rather than be convinced by others."* Participants mentioned how *"people tended to cluster into ideological groups,"* and stated that *"group formations seemed coherent with each member's affiliation."* Another discussed how they found it very believable that *"people would group up when views were similar; but not the same."*

3.11.2 RQ2: Does the measure of believability in the generated conversations vary across conversations?

The discussions were generated by varying two parameters in the generator: Group Size (Small and Medium) and Discussion Duration (Short, Medium). After every discussion was described (both textually, and graphically), participants were asked *"How believable was the change in the opinions of the conversationalists through the discussion rounds?"*

The results of this rating has been summarized in Fig. 3.10 below.

We ran the Friedman test to see if there were any differences in perceived believability be-

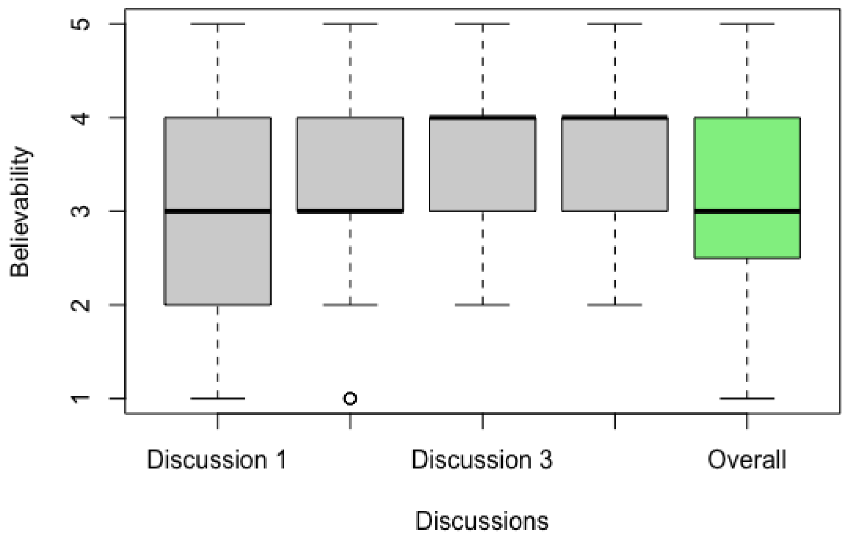


Figure 3.10: Four box plots showing the perceived believability rating across all 4 conversations as well as the overall believability.

tween the four discussions. We chose the Friedman test since we did not have independent observations among the 4 discussions analyzed since all survey participants analyzed all 4 conversations. We found there were no statistically significant differences in the perceived believability of the four conversations ($p > 0.05$). When asked what the least believable part of the conversation was, 4 of our 21 respondents mentioned they expected a more drastic shift in the opinions of the characters during the lengthier conversations, with one participant describing this as *"expected Mary's rightward shift to be a bit stronger (possibly getting to Moderately Right by Round 6)"* with another surprised that *"Shirley was not influenced by the other two in any way."*

3.11.3 RQ3: How similar is Lyra's clustering to how humans define and group like-minded NPCs?

For our discussion algorithm, we used Jenks Natural Breaks to group NPCs that expressed similar opinions to each other (Jenks 1967) and then evaluated for the goodness of variance fit (GVF) to select the optimum number of clusters. Survey participants were shown a chart depicting the opinions of the NPCs on our political scale, and asked (a) How many groups of like-minded conversationalists would form? (b) What groupings of like-minded conversationalists did they expect to see? Respondents used information about an NPC's opinion provided (both textually and depicted on our simplified political scale) to answer these questions.

Table 3.6: Describes respondents' agreement with Lyra's clustering results and the highest rated clusters.

Discussions	Model Clustering	Best Respondent Clustering
Discussion 1	0.1428	0.666
Discussion 2	0.5714	0.5714
Discussion 3	0	0.238 (tie for best cluster)
Discussion 4	0	0.333

For the second question, participants were free to choose from a list of groupings that the algorithm evaluated as the highest score for each possible value for number of clusters, or they could enter their own clustering if they disagreed with the choices given to them.

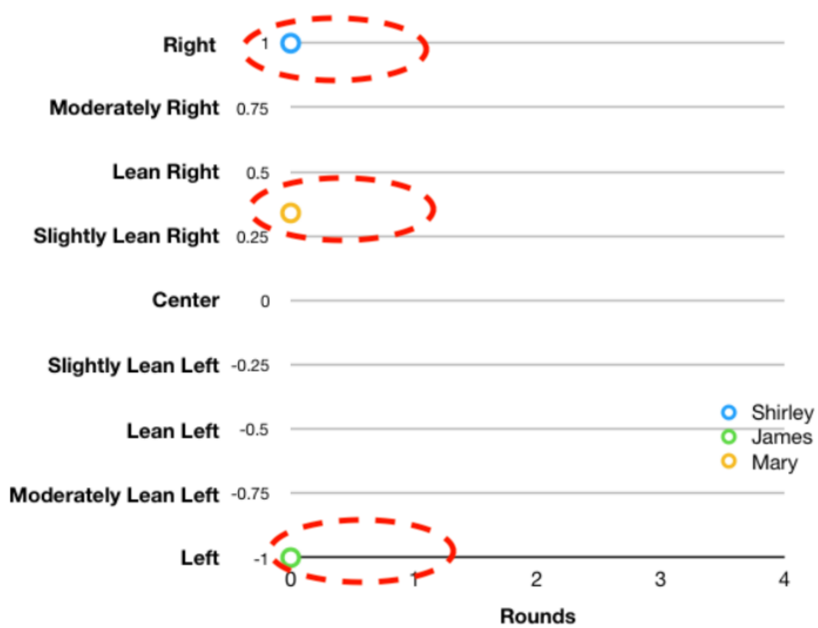


Figure 3.11: 57.14% of respondents agreed with our algorithm, and clustered the NPCs depicted here into 3 clusters (shown by the dashed red circles), one per NPC, with each NPC disagreeing with the other opinions proffered.

On the whole, only 27% of respondents agreed with the number of opinion clusters generated by our algorithm. Additionally, only 17.8% of respondents agreed with the choice of clustering made by our clustering algorithm. We have summarized the clustering agreement across discussions in Table 3.6. While the Jenks Natural Breaks Optimization algorithm tries to reduce the sum of the squared deviations from the cluster's mean, this optimization created a greater number of clusters than the numbers suggested by our participants 70.23% of the time. This can be seen in Fig. 3.12. The respondents chose to create their own clustering, with 50% of the total respondents in agreement on two similarly ranked alternatives. We have shown one of these groupings in Fig. 3.12. In contrast, our algorithm generated 7 clusters for the 9 NPCs during the start of the discussions. However, after the second round, our algorithm's clustering results agreed with that of the majority of the respondents. The change in the views of the NPCs during those rounds can be seen in Fig. 3.7 in the Study Design section of this paper. During the feedback for the conversations, survey takers talked about the clustering of NPCs into coalitions through the conversation favourably.

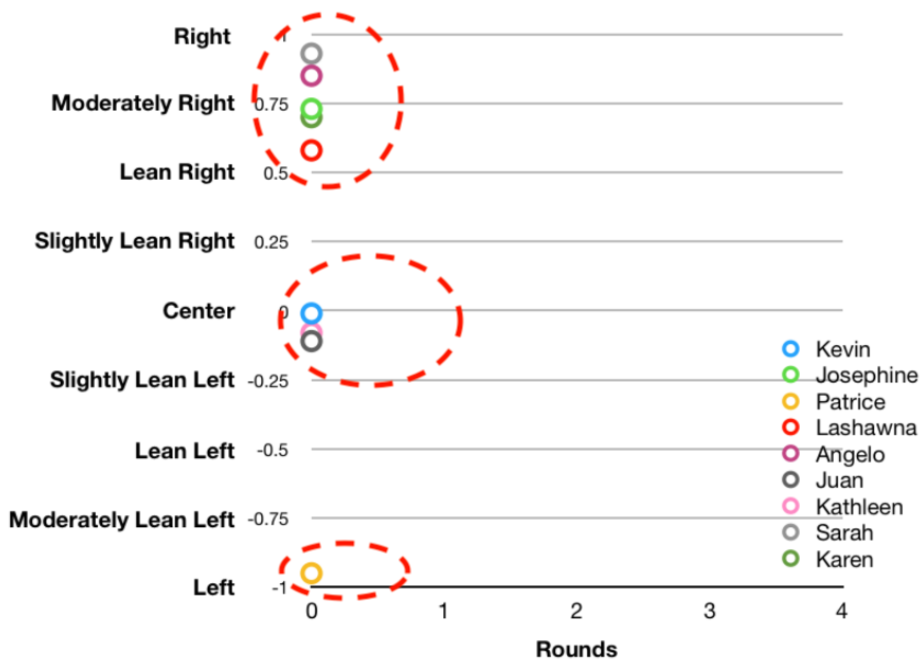


Figure 3.12: Respondents clustered the opinions depicted here into 3 clusters (shown by the dashed red circles).

3.11.4 RQ4: Does using the Lyra model impact the believability of the virtual characters?

Respondents were asked to rate how believable the change in the opinions of the conversationalists was through all the rounds of discussion for each conversation. They were provided with a Likert scale ranging from 1-Not believable at all, to 5-Very believable. Overall, the four conversations had a mean believability rating of 3.3 out of 5. We then qualitatively analyzed their open-text responses to our discussion questions. We report some of the more interesting responses and results from our qualitative analysis below.

Most Believable

When asked what the most believable part about the conversation was, respondents had varied responses. The most frequently mentioned themes from their responses have been summarized in Table 3.7 below.

Theme	**Frequency**
NPC mentioned unprompted	23
NPCs standing ground	18
Similar views converging	12
Influence from groups	10
Used political affiliation stereotype	9
Influence by an individual	8
Polarization	8

Table 3.7: Frequently occurring codes in response to what respondents found most believable

Breaking down the *#NPCMentionedUnprompted* tag further by the discussion we find that survey respondents interpreted the change in the characters views in the way our algorithm performs it finding that to be the most believable part of the conversation. For instance, in Discussion 1, noting an NPC sticking to their convictions (with a prevalent *#StandingGround* tag), *"Helga started at Left; moved to centrist and then closed at left."* In Discussion 1 and 4, we also see that influence exerted by other NPCs was accurately recognized (i.e. coded by the *#IndividualInfluence* tag) with statements such as *"Amy was swayed by William"* or *"Ada and Johnnie matching their views,"* being the most believable part of the conversations. Next, in Discussions 2 and 3 the *#GroupInfluence* exerted on NPCs was noted, with respondents pointing out it seemed as though an NPC changed their mind only because they seemed outnumbered or due to peer pressure stating, *"The fact that James had not changed drastically on his political opinion but has opened up his opinion to uncertainty seems believable since he is outnumbered in the group,"* or that *"Lashawna swaying slightly more conservative because she had a very convincing and large group and this would easily move her to similar opinion."* Additionally, respondents discussed the *#Polarization* of views in Discussion 1 with statements such as, *"That over time and rounds of arguments consensus develops around two poles of thought; even though within the poles there's a range of opinion/degree of certainty,"* and that *#SimilarViewsConverge* stating, *"No drastic changes in views but groups did come closer to same opinion on both sides."*

Finally, an interesting observation was that a lot of times people allowed their bias to affect their judgement and *#UsedPoliticalAffiliationStereotypes* while discussing the most believable part of the discussion summary they had just read. This despite the fact that our analysis of RQ1 found these biases did not affect their rating of the believability of the conversation. Respondents described how in Discussion 1 *"The consistency with which the Right Opined people stuck to their stand."* With Discussion 2, participants thought it was to be expected that *"that the most liberal person would be the person most open to changing their mind,"* while with Discussion 3 they found *"that the centrists didn't change their opinion much"* was very believable as was the fact that *"the five on the right [were] sticking together."* Similarly, with Discussion 4 they expected and found believable that *"Lefts found common ground and reached equilibrium."*

Least Believable

When asked what the least believable part about the conversation was, respondents had varied responses. The most frequently mentioned themes from their responses have been summarized in Table 3.8 below.

It is heartening to note that 5 of the 21 respondents found Discussion 2 to be entirely *#Believable* and could not describe the least believable part of the conversation stating, *"I find it believable"* as their responses.

Analysing Table 3.8, and the *#NPCMentionedUnprompted* tag, we initially find it was of almost double the frequency as it's an occurrence in Table 3.7 with 44 of the 68 responses in this section specifically calling out individual NPC behaviours as not believable or unexpected. Of Discussion 1, respondents discussed how they did not believe Helga should

Theme	Frequency
NPC mentioned unprompted	44
Changed Opinion	19
Decreasing Certainty	11
NPCs standing ground	10
Believable	6
Influenced by Article	6

Table 3.8: Frequently occurring codes in response to what respondents found least believable

not have been influenced by the article (i.e. *#ArticleInfluence*) as much as they were, further triggering the change in the views of Ashley and Ada. In Discussion 2, they brought up that they found it unbelievable that *"James (someone who was extreme left) was swayed by [the Centrist] article"* as much as they were. Two participants pointed out that it was the most believable and the least believable fact that Shirley *#StandingGround* was unbelievable stating, *"Shirley had no uncertainty in their views"* and *"not influenced by the other two in any way."* Of Discussion 3 modelled in Fig. 3.7, the *#ChangedOpinion* of Juan was pointed out as showing similarities to human conversations with one respondent stating that *"the unexpected move of Juan towards the Left and Patrice's position feels like the kind of strange turn that might happen in a real conversation - in a large enough conversation you will see some people's opinions change."*. However, this participant listed the same fact as both the most and least believable part of the conversation, wondering why Juan would change their views. Finally with Discussion 4, participants found Kenneth's *#ChangedOpinion* unbelievable, stating that they didn't think that *"Kenneth wasn't persuaded much at all; and shifting to the right seemed weird,"* or that *"William would be so persuasive [towards Kenneth] with such fluctuating levels of uncertainty"* was unexpected. While the *#DecreasingCertainty* tag was found to be frequent, there was no consensus amongst respondents on how this affected the least believable part of the discussions with the tag occurrence being sparsely distributed.

Reasoning Queries

When asked to reason about an NPCs change in opinion, certainty or attitudes, respondents had varied responses. The most frequently mentioned themes from their responses have been summarized in Table 3.9 below.

Theme	Frequency
Individual Influence	19
NPC mentioned unprompted	15
Opinion Attitude Difference	12
Infer Facts not provided	11
Group Influence	10
Certainty Convinces	10
Lacking Support	8
Emotions Attributed	7

Table 3.9: Frequently occurring codes in response to the reasoning questions asked

#IndividualInfluence by a fellow conversationalist NPC was denoted as the major factor influencing the change in the uncertainty of Ada (in Discussion 1) and Amy (in Discussion 4), with quotes such as *"She was uncertain to begin with and her groupmate; who was the most knowledgeable (ie if no of prior articles read is an indicator of knowledge); was also wavering her convictions,"* or *"The influence of William's arguments [swayed her]."* 7 of 21 respondents blamed William (*#NPCMentionedUnprompted*) in Discussion 4 for Amy's uncertainty with statements such as, *"I think they were aware of their drift in position and how convinced they were by William's arguments."*

Respondents in Discussion 3 concurred that it was an awareness of an *#OpinionAttitudeDifference* that caused Juan's change in opinion. They quoted, *"He didn't want to seem biased externally so wanted to be portrayed as a centrist; but was privately left-leaning,"* and concurring that *"their view was probably more left-leaning than they initially realized."*

Additionally, for Discussions 1 and 2, both discussions with a smaller group of conversationalists, respondents pointed out that the certainty of the other NPCs helped sway opinions (i.e. *#CertaintyConvinces* stating, *"You must assume this is because of Johnnie's certainty"* or *"The opposition members confidence and articulation was strong."* We believe this smaller number of conversationalists is what influenced both discussions' NPCs to be tagged as *#LackingSupport*. Respondents discussed NPCs having the *"feeling of being marginalized,"* and that they seemed to be a *"lack of support from like-minded people."*

In discussions 2 and 4, both discussions of a longer duration, respondents pointed out that *#GroupInfluence* was a factor in changing the NPCs views. Respondents stated how the *"opposition had convincing arguments or [that there was a] tendency to want to agree with the majority,"* and that there was a tendency for an NPC to cave on their views since they would associate them with *"temporary bias because of peer-pressure in a group of majority* *conflicting opinions."* Interestingly, in discussions of shorter duration (i.e. Discussions 1 and 3) respondents were more likely to *#InferFactsFrom* the study that was not initially provided to them. They made statements about how the NPC must *"support for innovation and reform strongly"*, or seemed to *"value [the] Rights and Interests"* of the other conversationalists more. These conversations also tended to have stronger *#EmotionsAttributed* to the constituent NPCs with respondents attempting to articulate the emotional distress of the conversationalists saying, *"Changing one's political identity on an issue isn't an easy task and can result in much internal conflict and therefore high uncertainty"* or blaming the *"feeling of being marginalized"*, or that an NPCs *"competitiveness seemed to be declining"* or that an NPC didn't seem to *"care for the well-being"* of the rest of the population.

3.12 Discussion

With this section, we revisit the goals of our project and discuss each. We also discuss our major findings, along with our future work plans.

3.12.1 G1: To generate descriptions of the change in the opinions of the conversationalist NPCs that allowed readers to follow the NPC's reasoning

Our design process for these generated conversations as described in our section, *Designing Legible Simulation Output*. Of the 21 respondents, 17 were able to interpret the conversations and use them to reason about NPC behaviour. 4 participants stated that they had difficulty following the conversation description. One participant mentioned that the descriptive text provided by us made it *"difficult to align with [their] own mental model of the dynamic. The graphs help; but the textual description is pretty poor [and] too abstract."* Overall, we believe that these responses satisfy our goal. Our system can produce modifiers and keywords that state the intention of the characters in a manner that meets the expectations and match the mental model of the reader. In the future, these could be used to produce natural language dialogue utterances.

3.12.2 G2: To evaluate these generated conversations with a human subject study for their believability

We describe the design and method of our study in the *Study Design* section. One limitation of our study was the small number of respondents and the fact that they were mostly on the left of the political spectrum. Our population sample was not normally distributed, making it difficult to test for statistical significance in our analysis. Overall, the four conversations had a mean believability rating of 3.3.

3.12.3 G3: To extract insights from the study to inform future research

We see these insights in our section on *Analysis of Results*. With our four research questions we conducted a summative evaluation of our simulation. With our qualitative analysis, we learned how respondents felt NPCs believably form coalitions. Our reasoning questions showed that most respondents were able to interpret and expect the change in NPC opinions in the way our algorithm performed it. An interesting point to note is that readers expected NPCs to stand ground and not change their mind in many cases, claiming that this added to the believability of the discussions. Additionally, some respondents displayed emotional responses to the conversations they read (for instance, stating that they found it *"believable but depressing that [none of the NPCs] ultimately changed their minds [on Immigration] at the end of Round 3"*), while others attributed emotions to the NPCs involved discussing NPC competitiveness, or caring for the well-being of the population, or the NPCs support for reform.

Togelius et al. (Togelius et al. 2013) discuss how game believability is a critical subcomponent of the player experience. It can be linked to a stream of player emotions triggered by events occurring during interaction but also related to cognitive and behavioural process during gameplay. They continue to describe how games with believable elements can elicit emotions in the player. Additionally, several authors argue that the appearance of human intelligence or human-likeness adds value to a computer-controlled character and thus to the quality of gameplay (Togelius et al. 2013; Champandard 2003; Bateman and Boon 2005). We believe that evidence described above of these emotional responses elicited in the player, and the emotions and humanity ascribed to our NPCs can be taken as further evidence of the believability of our system. We suggest this shows that despite the simplicity of our chosen discussion template readers are primed to imagine complex layers of interactions between the NPCs. Our simulation was able to invite users to use their imagination and provide to them a more immersive and compelling narrative effect.

3.12.4 Modeling Social Influence and Simulation

In their responses, respondents pointed out some interesting features of social dynamics that we did not intentionally simulate, attributing changes to these social phenomena. One of those was the existence of an Overton Window in some of the discussions. They pointed out when an NPC changed their mind *"because she was an outlier, and had the most extreme view"* or in another case how *"everyone else expressed a more rightward view; making Ashley's view appear more extreme left that it actually was."*

The participants also pointed out when *#Polarization* seemed to be occurring with the groups clustering away from the centre. One participant stated that *"no substantial agreement was reached; which is what you might expect from an argument where people's views start out very highly separated from each other,"* while another pointed out that this type of polarization could lead to the feeling that NPCs were *#LackingSupport*, feeling marginalized or as though they were outliers with the participants tending to *"cluster away from centrism."*

Most interestingly, with our analysis of the most believable part of the conversation, we noticed an interesting pattern of how readers discussed *#GroupInfluence*, *#Polarization* and how *#SimilarViewsConverge*. Survey respondents spoke of this as a matter of fact, pointing out how individual members ceded to peer pressure or conformed stating they were *"outnumbered"*, or that an NPC *"was in the minority so probably felt uncertain"*, or how *"deliberation within a group is important and with the right convincing you can change someone's mind"* and that *"there is some power in group mentality"*. We believe these observed patterns further strengthen and support our hypothesis on social rules. Beliefs and attitudes that go against a group's values are looked down upon unfavourably by the members of the group. This would allow with our approach, for the group and cultural rules to emerge organically through the course of interaction with their members.

While these findings were unexpected, we are heartened that our simulation can model and generate these recognizable social phenomena. We believe this further adds to the immersion and believability of the characters.

3.13 Simulation Pains and Gains

The original difficulty with setting up Lyra was, as I have discussed at the start of this chapter, reinventing the wheel. In other words, having spent a couple of months trying and yet failing to modify existing simulations such as Talk of the Town to include my Discussion model, I had to give up and start modeling Lyra from scratch. This process took time, I had to design and set up social interactions as basic as movement tracking, and enrolling in school, and interactions as complex as ensuring a child grows up with biases from their parents (a modified genetic algorithm with some mutations thrown in), and social relationships and power dynamics. As a new PhD student, I spent a large amount of time trying to understand how to model the innate characteristics of my simulated humans as well as some believable social interactions.

Having spent a considerable time rediscovering and reimplementing the basics of NPC simulation that those before me had discovered, I was still left with one huge problem. During the evaluation phase, how would I know whether the characters seemed more believable because of my newly added, distinct discussion model? It was possible that the characters were more believable due to the way I designed, implemented or represented a social interaction or a social relationship. Finally, we made the choice that we would only test the generated conversations for believability. Thus, our human subjects were not made aware of any of our simulated power dynamics or other influences characters had on one another that swayed their mind. Instead, they were only aware of the final outputted discussion or conversation as shown in Fig. 3.7 without being told, for instance, whether one of the conversants was a professor who had less uncertainty in their own views, and were thus unlikely to change them.

Thus, we found our evaluation to remain limited. Further discussions on the subject post our paper acceptance and talk led us to imagine a scale or taxonomy that indicated the granularity of interactions, or a dictionary of interactions that could be selected, while plugging on or off the discussion model to evaluate it's impact on the character's believability. It was this thought however, lead to the construction and design of our eventual taxonomy.

CHAPTER

4

ANALYSIS OF EXISTING SOCIAL SIMULATIONS

This chapter overiews Research Thrust #1 *(completed)* that addresses my first research question (RQ1). The first thrust of our project was to understand and organize knowledge concerning existing design processes and methodologies involved in the development of social simulations.

RQ 1: *How do multi-agent social simulations in the narrative intelligence domain currently model social inter-agent behaviors?*

To answer RQ1, I conducted an in-depth analysis of a selection of research projects and commercially available popular social simulation game franchises. I collated existing knowledge concerning agents modelled by these simulations. Furthermore, I was able to identify a lack of consensus and inconsistencies in the vocabulary used to define and describe research advances. This lack of consensus directed my inquiry and analysis of existing research projects. I asked myself:

RQ 1.1 What, if any, are the key barriers to the process of researching or designing social simulations that can be identified from a survey of their accompanying literature and code repositories?

This chapter answers my research question above. I overview the underlying factors for my choice of research artefacts to overview as my representative data set. I describe my review process for identifying and coding discovered inter-agent behaviors. Next, I describe the process of analyzing, categorizing, and organizing the codes to create the taxonomy. Finally, I discuss my findings, and answer my sub-question, regarding the barriers to the process of researching social simulations in Section 4.3.

4.1 Choice of Artifacts for Analysis

Over the past few decades, agent-based modelling techniques and social simulation have been increasingly popular in the field of entertainment intelligence. Social agents have been used to study emergent narratives (Ryan et al. 2015; Kreminski et al. 2019), interactive drama and improv systems (Eger and Mathewson 2018; Magerko et al. 2009), model characters for players to interact with (Mateas and Stern 2003; Ware et al. 2019), and to study macro-level phenomenon produced by changing micro-agent details and interactions (Porteous et al. 2010).

Our goal was to conduct a systematic review comparing results from research-based and commercial social simulation works in the entertainment domain, allowing us to categorize an initial set of social interactions.

4.1.1 Choice of Research Artifacts

Simply perusing the list above, one can see a wide range of research artefacts to choose from. We chose to limit ourselves to social agents and social simulations as defined by prior work (Brassel et al. 1997; Wooldridge and Jennings 1994; Moulin and Brahim 1996). To select the research artefacts to review, we systematically reviewed past submissions on databases such as AAAI, ACM, and IEEE. We looked at submissions to AIIDE, GDC, FDG and the Journal of Artificial Societies and Virtual Worlds. Additionally, we looked at Google Scholar, and ResearchGate to identify work similar to the projects and papers we chose to review. We searched for a range of topics, with keywords including but not limited to social agents, social simulations, multi-agent social simulations, non-player characters, agent interactions, agent models, virtual characters. We collated all papers, and code published or publicly available for each shortlisted project. We shortlisted projects that had discoverable published papers and code and the granularity of details regarding the agent interactions in the simulation that we were looking for. We reached out to authors where necessary for more details or clarifications.

In summary, we looked specifically at social agents (Wooldridge and Jennings 1994; Moulin and Brahim 1996). We decided to constrain ourselves to analyze the differences in large-multi-agent systems (Brassel et al. 1997) with well-detailed environments, both common and specific. We searched multiple databases for a wide range of search terms and related terms described above. In the end, we were able to conduct an in-depth review of 7 academic research projects: Islanders (Ryan 2016), CiF/Prom Week (McCoy et al. 2011a), TALE-SPIN (Meehan 1977), Lyra (Azad and Martens 2019), Thespian/PsychSim (Si et al. 2005; Pynadath and Marsella 2005), Talk of the Town (Ryan et al. 2015), and Versu (Evans and Short 2014).

4.1.2 Choice of Commercial Games

We decided to include two social simulation video games to analyze the gaps between industry practice and research. In lieu of the easy availability of code artefacts, we looked at existing official and community wikis for the games. The rest of the criteria for the game selection remained the same as those of the research projects. We also searched for popular, topical games (Stieg 2020; Gray 2020), shortlisting the selection by the number of players globally.

We narrowed our choice down to two commercially available game franchises, *The Sims* (Maxis 2003), and *Animal Crossing* (Nintendo 2001), with their rich community wiki information. We feel these choices are both current and topical within our community.

4.2 Review Process

Once we had shortlisted our eight academic research projects and two commercially available games, we could begin our review for the same.

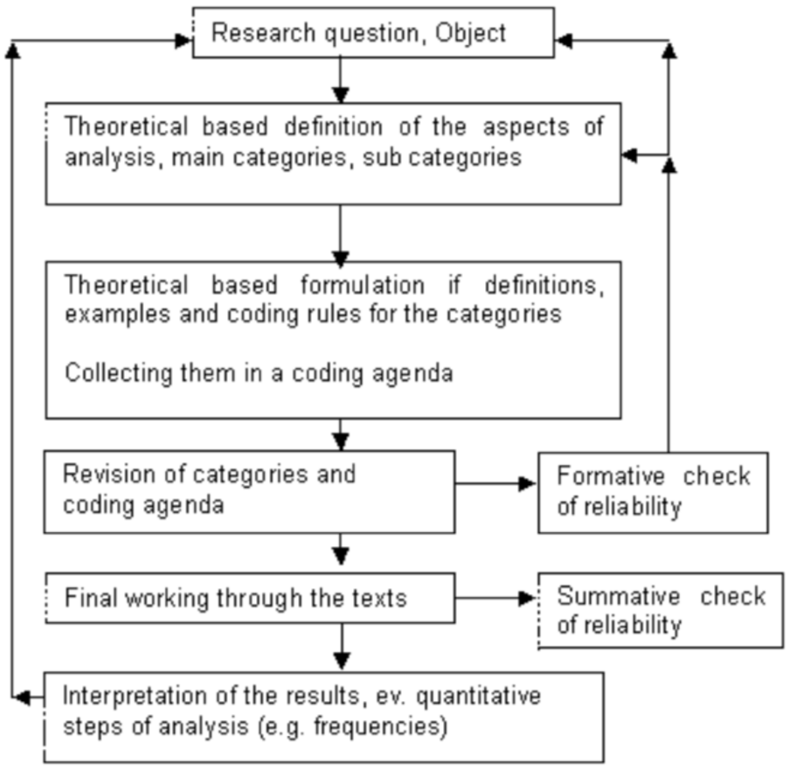


Figure 4.1: Step model of deductive category application (Mayring 2004) used during the Review Process.

e

4.2.1 Identifying Interactions

We define *interactions* to be any interaction between the social characters in the studied simulations. This includes their actions and social behaviours, emotions, and ability to form relationships. Since we were interested in social interactions, for the purpose of this survey, we excluded individual interactions such as sitting, standing, moving between locations, or fishing when alone.

We first collected all available written material on the selected projects. This involved collecting and combing through any papers or wikis available for each project during various development stages. When available, we cross-listed interactions mentioned within academic papers with their code artefacts. In every case, since the written publications had disparate goals (ranging from entertainment to simulation of social science phenomenon), we found the code artefacts and repositories artefacts contained a more complete set of interactions than mentioned in the paper. We listed each of the interactions found this way.

We were able to discard some interactions from *The Sims* as too arcane or specific for our purposes. For instance, discarded interactions include a Vampire screeching "Bleh!" to a subject of their choice, or a Sim asking a High Evil Witch to, "Teach [them] the Path of Darkness." These interactions were marked as *Special* even within the wiki in question.

By the end of this stage, we had a list of around 700 interactions across the simulations. Additionally, we had metadata around the interactions pertaining to when they could be triggered, and the relationship type required, social events, moods or emotions they would induce.

4.2.2 Deductive Category Development

We performed an initial open coding and used content analysis to analyze the curated list of interactions. As we were familiar with the underlying theory in virtual character interactions, our initial codes had a structured and directed approach (Hickey and Kipping 1996) described by Mayring (2004) as deductive category development application. Our analysis process followed Mayring (2004)'s step model as depicted in Fig. 4.1. The data was read from start to end to obtain a sense of the whole. We noted down first impressions, and thoughts to capture key concepts based on the variables of interest (Potter and Levine-Donnerstein 1999). Next, we tackled each group individually, making notes for an initial coding schema.

Table 4.1: Initial categorization during an early iteration of the Mayring Deductive Development process (Mayring 2004).

Category	Initial Subcategories	Example Interactions
Relationship	Friendship, Enmity, Romance, Vocational (Classmate, Colleague)	Get divorced (Romance)
Mood or emotion	Happy, Gratitude, Embarrassed, Angry	Take an angry poop (Angry)
Personality	Flirty, Angry, Competitive, Friendly	Conversation flirt (Flirty)
Type of Interaction	Social, Individual, Normal, Romantic	Declare an enemy (Social)

The first few iterations through our dataset produced the following general categories and subcategories (Mayring 2004), as shown in Table 4.1. For instance, the "Get Divorced" interaction was coded under the Romance subcategory since it directly affected the romantic relationship between two characters. These categories helped us sort through the somewhat unwieldy set of interactions.

Performing open coding analysis, we read through each of the collected interactions to note down codes to describe them (Miles et al. 1994; Morgan 1993). Many interactions in the dataset were found to have a valence or affinity associated with them. For instance, "Declare an Enemy" had a negative influence on the friendship relationship, or "Give a backrub" was found to influence the romantic relationship. These interactions were found to have effects that were encoded into the projects. In the research code bases, we found that projects had explicit reference to these as "Sparking Romantic relationship", or "Charging social meter" (Ryan et al. 2015). Similarly, these effects were documented in the wiki of the commercial games (Maxis 2003). In Table 4.2, we detail the effect for the "Admire" interaction on the daily and lifetime relationship between the initiating (character A) and receiving (character B) Sims based on whether the interaction is accepted or not from *The Sims* wiki (Electronic Arts 2000).

Table 4.2: Effects of the "Admire" interaction on the various relationships between two virtual characters depending on the acceptance of the "admire" interaction in *The Sims* (Electronic Arts 2000)

Interaction: A(Admires, B)	**Accept**	**Reject**
A's Daily Relationship with B	5	-10
A's Lifetime Relationship with B	1	-1
B's Daily Relationship with A	4	-7
B's Lifetime Relationship with A	2	-2

To account for these valence shifts in relationships or behaviours, we created a separate Valence category with tags such as "Increasing", "Decreasing", and (in cases where a change was identified but not the direction) "Influence". These tags could be used in combination with the others to indicate the expected outcomes of an interaction. For instance, #Romance#Increase indicates that interaction could produce an increase in the strength of the romantic relationship between the initiating and receiving character.

We were able to list an initial set of 54 tags that could be used to tag each interaction. These codes were organized into meaningful clusters to make lookup easier for the evaluators of the coding schema. We were also able to list the explicit definitions, examples, and coding rules for each deductive category discovered through iterative analysis. A few codes

Table 4.3: Some high frequency categories and subcategories with corresponding definitions after the deductive category development application

Category	Subcategory	Definition and Example Interaction
Social Interaction - occurs between multiple characters, affects relationships	Admonish	To reprimand another. Eg. Insult, Patronize, Punish
	Appreciate	To recognize the worth of, cherish, or praise another. Eg. Give gift, make positive utterance
	Entertain	To provide another character with amusement or happiness. Eg. Prank, Offer drink
Communication - exchange of information, or feelings. May be added to knowledge base.	Verbal	Relating to or in the form of speech or verbs. Eg. conversation flirt, announce promotion
	Gesture	A physical movement to express an idea, or meaning. Eg. friendly hug, give medicine
	Physical	Perceived to be or have an affect on a tangible, sensation (as opposed to verbal, or emotional). Eg. embrace, commit murder
Relationship - interactions that change the status or valence of a relationship	Familial	Requires a familial relationship or changes the status or valence of the familial relationship. Eg. Get married, have baby
	Romantic	Requires a romantic relationship or changes the status, or valence of the same. Eg. physical flirt, break up
Change in Valence - an increase or decrease associated with the valence (or strength) of the interaction	Increase Decrease Influence	Induces or influences a change in valence. Should be used in conjunction with another code as a modifier. Eg. #SocialRank#Influence

from this initial coding scheme (or codebook) have been described in Table 4.3.

4.2.3 Validating the Coding Schema:

To further establish rigour, reliability, reduce the coding scheme's discriminant capability, and validate this initial codebook, two additional independent coders were recruited. The coders would be categorizing a random 15% of the shortlisted dataset of virtual character interactions.

The new coders were unfamiliar with the project and the eventual goal of developing a taxonomy. They were given a brief description of the research to explain its purpose. We then walked through the definitions, codes in our codebook, and one defining example of each code selected from the remainder 85% of our interactions dataset. After an initial discussion and walkthrough, the coders initially had a few questions about existing definitions and rules in the codebook. The codebook was updated further to clarify the same to the coders.

The coders were instructed to use as many codes or categories for an interaction as they felt necessary. They were also instructed that they could create new categories if they felt the existing coding schema could not describe the social behaviour or interaction correctly. One of the authors also re-coded the same section of interactions as the coders. The coded sections were compared and negotiated. Some adjustments were made to the codebook during this phase. For instance, the tags #Relationship:Friend and #Relationship:Enemy was found to be confusing. There was no agreement as to what interactions could cause a change in the status from friend to enemy (except, the interaction "Plot Murder"). Instead, these were replaced with a Relationship:Social tag and associated with a valence, for instance, Relationship:Social:Increasing denoting the social relationship would increase between the characters due to the interaction. The tag #SocialNetwork was added to indicate an increase in the social network of an NPC (for instance, in response to interactions such as "Hire Employees" or "Establish Town"). Depending on an individual NPC's threshold for social relationships, other relationships could evolve from the character's social network over time, for instance with characters gaining friends, romantic interests or enemies. A few more code definitions and examples were clarified and refined based on feedback from the coders.

Next, we calculated the inter-rater reliability between the two coders and the author. Since multiple codes could be applied to each interaction (a one-to-many coding scheme), we used the Fuzzy-kappa statistic (Kirilenko and Stepchenkova 2016), an inter-rater reliability statistic based on Cohen's Kappa and modified for the application of multiple codes to a single response. Our calculated values of inter-rater agreement (82.86%) and the inter-rater reliability (0.819, Fuzzy Kappa) indicate an "excellent" to "almost perfect" consistency for the application of the model characteristics coding scheme (Kirilenko and Stepchenkova 2016; McHugh 2012). The high agreement allowed us to finalize our coding schema. We proceeded to use these codes to tag the rest of our interactions data set. Once tagged, we were able to analyze our coded dataset. This analysis helped us develop our taxonomy, and understand more in-depth the issues faced by social simulation researchers and developers.

4.3 A Simulator's Predicament

Our taxonomy-building work arises from observing several pain points in our research and implementation practice, as well as in the surveyed social simulation projects. In this section, we describe the pain points discovered from our in-depth survey in detail and discuss how they motivate our research goals.

4.3.1 What do words mean?

In prior literature (Treanor et al. 2015; Shapiro et al. 2015; McCoy et al. 2012; Samuel et al. 2015), a *social state* refers to any subset of emotions (e.g. happy, sad), moods (e.g. good mood, negative mood), relationships or intentions of the virtual characters, or personalities (e.g. using the OCEAN model). Social states occur either in individual characters or are shared between characters associated through affiliations across regions, occupations and can affect the behaviour of an agent.

Figure 4.2: Example of how our surveyed artifacts implement a single social behaviour, "brag". From the left, we see the interaction in *Prom Week* (McCoy et al. 2011a), *The Sims* (Maxis 2003) and *Animal Crossing* (Nintendo EAD 2020) artifacts.

Prior literature has shown that *social relationships* are instrumental in driving narratives and social behaviours (Ho 1998; Meehan 1977). However, models of relationships differ across simulations. The start of a relationship may compare personality models or randomly generate a "spark" amongst the characters (Ryan et al. 2015). Another simulation may create relationships based on group attachments and affiliations (Azad and Martens 2019). Still others look at forming relationships based on resources exchanged or traded such as love, status, information, money, goods, or services (Nintendo EAD 2020; Weiss 1991).

A *social verb or action* may vary in granularity across models. Two agents collaborating on a task may be considered social (Castelfranchi et al. 1997), and two agents simulating a family over a decade are also social. Similarly, there is no clear classification of the different ways in which agents can communicate: each model proposes its own interaction structure. For instance, a model may choose to simulate an agent being “happy” as a binary value if something good occurs. However, another system may simulate happiness using a more complex emotion model taking into account intensity, duration of happiness, and its intersection with other emotions experienced at the time (Lisetti 2002; Pynadath and Marsella 2005; Si et al. 2005). Without a shared vocabulary or expectations, it is difficult to compare or discuss these design choices.

Finally, what “social interactions” are considered meaningful to form or influence *social relationships*? The projects all influenced the relationship amongst virtual characters, with the assumption that performing “social activities” would improve their relationship, and “anti-social” behaviour would degrade it. The natural question that arises would be should all such social activities be treated the same? How much should each activity influence a relationship?

4.3.2 Reusing Social Models

Existing social interaction models have varied contexts, rules, narratives and even timescales for the interactions in their NPC simulations. Social simulation systems embed differing assumptions within their language of interactions that makes it hard to compare them. Fig. 4.2 shows how the surveyed artefacts implement a single "brag" interaction differently using a combination of verbal, or gestural cues. The lack of consensus in what comprises a social interaction, and what features or requirements must be developed to simulate a social character successfully leads to research groups frustrated and with no choice but to reinvent the wheel, creating new simulations from scratch.

This behaviour has generated a vast quantity of prior work on NPC models with the potential to lead to reuse of accepted behaviours, constraints, or “social practice templates” and norms instead of having to resort to reinventing the wheel. However, it is currently unclear how to go about the same. A key pain point that many new researchers in the field experience is that in order to test any new interactive technique or AI behaviour, they *must first reinvent the wheel*. For instance, in their project *Talk of the Town*, Ryan et al. describe creating a rich society of characters within a small American town (Ryan et al. 2015). Similarly, in *Lyra* (Azad and Martens 2019) to test their discussion model Azad and Martens had first to implement a believable society from scratch. In both cases, authors made independent decisions on the programmatic encoding of similar social norms and interactions, even though the projects simulated the same granularity level.

This leads to a large proliferation of implementations, each with a varying subset of the human life simulated, and variation within each subset of simulation and implementation. This is because current simulations make it hard to reuse existing models, and can hamper progress.

4.3.3 Reproducibility and Evaluation

Far too often, it is difficult to understand the details of the implementation of NPC models. Richiardi et al. discuss how current models have difficulty with replication and identify the aspects of models that affect replicability, namely programming language, tools, representation formalisms, and development methodologies (Richiardi et al. 2006). Many well-accepted models communicate conceptual descriptions excellently, but do not define their models using unambiguous formal notations that could be reimplemented in a new project.

We posit that having a common taxonomy, and a vocabulary with set expectations of implementation details when choosing to describe projects will improve reproducibility and evaluation. Game designers can actively understand and consider existing project features, as well as the novelty, and their contributions when discussing their work. A taxonomy and vocabulary would be the right step towards a formal definition simulating social human behaviours, and interactions.

4.3.4 Comparison of Models

Traditional practices in social sciences rely on a very well established and implicit methodological protocol, both with respect to the way models are presented and the kinds of analysis performed. The field of machine learning also has very precisely defined benchmark tests and metrics commonly accepted as a useful measure of model performance or evaluation (with Code 2021; Repository 2021). Models created for social character simulation lack such a reference to an accepted methodological standard.

Additionally, it is currently difficult to compare the assumptions made regarding the input and output of the social simulations analyzed in this paper. For instance, *Lyra* (Azad and Martens 2019), *Juke Joint* (Ryan et al. 2016a), and *Prom Week* (McCoy et al. 2011a) have social models that can influence the views of a character on various topics such as politics, a personal dilemma, or social status and popularity. These models can affect change in the relationships of the interacting characters. However, the streaming output from the simulations themselves can not currently be compared. In another instance, with Table 4.2, we see the documented results of a single "Appreciate" interaction being performed in *The Sims*, and its effect on the *valence* of the relationship between the participating characters (Maxis 2003). Similarly, the attack interaction reduces the valence of the relationship by a factor of -60 to -100. Similar interactions exist in *Prom Week* (McCoy et al. 2011a), *PsychSim* (Pynadath and Marsella 2005), and several other social simulations. However, currently, it is nearly impossible to compare their effects.

We set forth that there needs to be a more rigorous model-to-model analysis. Demonstrating social behaviours that are agreed upon by an existing body of research is interesting. However, we can learn a great deal from when these models do not agree. This would allow us to understand better and evaluate existing models, and replicate their behaviours.

4.3.5 Research Collaboration

Currently, social simulation research groups tend to work in isolation, with each group creating a new social simulation, framework, or engine from scratch. As discussed in Section 1.1, this leads to a proliferation of disparate implementation of frameworks, each aiming to test a single set of computational social science theories, or cognitive and social agent models. While each disparate model, thus created, has a wealth of knowledge associated with it stemming from the researchers attempt to understand and emulate human interactions, creating these models from scratch takes time and effort. Researchers aiming to study, for instance, a new model of conversation must first familiarize themselves with how to create a rich, society of characters, which relationships to model, and decide whether their choice of relationships modelled, or the power dynamics socially associated impact their knowledge model. The initial time and effort associated with the creation of a model can be reduced through research collaboration.

However, the existing disparate implementations are generally seen to have the agent, behaviours or interaction modes, and environments (Brassel et al. 1997) that define the simulations tightly coupled in the simulation engine. This tight coupling makes it nearly impossible for other users or research groups to experiment with alternative levels of authorial control, agent representations, personalities, decision-making logic, interaction modes or narrative settings without a major rewrite of the simulation engine logic and design.

Instead, one must incorporate the ad-hoc assumptions made by the simulation's original author about the agent model, interactions available, or environment into their own experiments (Johnson-Bey et al. 2022b). This lack of modularity in turn leads to authors opting to reinvent the wheel, leading to more disparate implementations available, and thus forming an inconvenient feedback loop that further reduces the reuse and reproduction of existing published models as well as the potential for research collaboration and hampering progress.

4.4 Developing the taxonomy

Our taxonomy was created as a result of three separate levels of analyses done on the surveyed projects.

First, a qualitative thematic approach using deductive category development on the coded interactions is described in the methodology section (in Section 4.2). Once the coding schema was finalized, we reviewed the codes to identify, organize, analyze and report the recurring themes discovered across the 700 interactions dataset. We used reflexive thematic analysis (Braun and Clarke 2006), first mapping and grouping initial codes that identified broader patterns, then gradually refining these categories, and testing them against our dataset. We were then able to review the derived themes, splitting and combining categories until we were satisfied with how they represented and told a story about our data. We named the themes, formalized their definitions and the rules associated with their use, and classified the sub-themes and interaction tags associated with each.

Once these themes were obtained, we did a second iterative analysis, deep-diving into the projects' code repositories (when available) and the published manuscripts associated. We searched both of these artefacts for the themes we had identified, trying to understand the underlying theory, ideas or concepts of the interactions comprising the theme, and reaching out to the authors when necessary for further clarification. By doing this, we understood how the projects encoded notions of social norms, rules, and theories into their social characters to make them more believable or relatable for their human players. At this stage, we also read player reviews, and play-through narrative accounts of their experiences with the simulations (where available) to understand the reception of the design choices. This allowed us to refine the rules and constraints associated with the identified themes.

Finally, to verify the taxonomy generated, I performed a graph analysis with using a graph-based database, Neo4j (Neo4j 2012). I was able to plot the co-occurrence and

A	F	G
From Tag - To Tag (Graph Analysis)	**intersection**	**similarity (low < 0.6)**
"#Envy"--"#Appreciate#Decrease"	**1**	**1**
"#Admonish"--"#Communication"	**9**	**0.9**
"#Angry"--"#Communication"	**2**	**0.6666666667**
"#Appreciate"--"#Communication"	**21**	**0.65625**
"#Appreciate#Decrease"--"#Communica	1	0.5
"#Brag"--"#Communication"	**4**	**1**

Figure 4.3: Early analysis using Neo4j to evaluate the taxonomy themes. In this example, the tag #Brag and #Communication co-occurred 4 times. The overlap coefficient (or similarity) was 1, which indicates every bragging interaction was seen as a method of communication between characters. In comparison, not every interaction leading to a decrease in appreciation was seen to be a communication type interaction.

similarity between the coded tags (as can be seen in Fig. 4.3)

4.4.1 Vocabulary and Nomenclature Decisions

The taxonomy's nomenclature was decided on after the themes, sub-themes, rules, and constraints for our taxonomy were refined and detailed. During this stage, we referenced project publication materials and community wikis, social science publications, and other existing vocabularies.

- *Project Publications and Wikis:* We looked at publications from the authors of the project detailing the theme we were categorizing. We made a note of individual project vocabularies, comparing similarities and differences. An example of nomenclature adopted in this manner is the attribute of *Relationship Symmetry* that matches the description in Evans and Short (2014)'s Versu.

- *Social Science Publications:* For most academic projects, we were able to identify the social phenomenon the project was attempting to encode from their associated published manuscripts. We were able to follow their reasoning and choices back to the social science, cognitive science or computational narrative intelligence sources they referred to. Additionally, we searched Google Scholar for existing nomenclature representing the phenomenon or rules we were describing when necessary. An example of vocabulary adopted from social science papers includes the *Principles* *of Contingency, Inertia, Regulation, and Interaction* that help categorize Emotion interactions in our taxonomy. These principles are widely accepted and discussed in further detail in social science papers (Kuppens and Verduyn 2017; Russell 2003).

- Other Vocabularies: We also used terminology from the logic, mathematical, and computer science domains where necessary if the earlier methods did not result in the development of a satisfactory term to express our findings. An example of adopted vocabulary that falls into this category includes the terms *Cardinality, Periodicity, and Exclusivity* listed under relationships in our taxonomy.

CHAPTER

5

THE LITTLE COMPUTER PEOPLE TAXONOMY

To resolve the inconsistencies and lack of consensus discovered in Section 4.3, I organized the knowledge collected for RQ1 to design a common taxonomy and rule-set for modeling social agents. In this chapter, I describe my second research thrust *(completed)* that addresses my second research question (RQ2).

RQ 2: *How can we consolidate the differences and similarities currently modelled in the social simulation agents?*

I describe the final taxonomy that resulted from our in-depth analysis of a selection of projects (both research and commercial). This thrust answered following sub-questions:

RQ 2.1: What are the overarching themes of social interactions between multiple agents that can be discovered through analyses of existing multi-agent social simulations? How do we consolidate the differences and similarities across the surveyed social simulations into a taxonomy within these themes?

RQ 2.2: Does the consolidation of the differences and similarities across multi-agent social simulations into a taxonomy enable us to:

- Identify the breadth, and depth of artificial intelligence, social science, and cognitive science narrative research explored.
- Identify unexplored territory in the space of social simulation design that could lead to exciting future crossovers between the social sciences, game design, and artificial intelligence.

I outline the themes discovered in existing social simulation works, and answer RQ 2.1 in the sections below. I collect and describe the results from RQ 2.2 into Chapter 7 where I hope to add to it further results from our expert evaluations.

5.1 Theme: Communication

Almost a third of the interactions discovered across the artefacts surveyed were tagged as *communication*. This can be explained by the research trend of designing intentional characters. As such, a large number of interactions are designed to communicate or demonstrate a specific intent of the characters involved. For instance, throwing a drink in another's face may be seen as a character strongly disagreeing with another or responding to another character's communicative interaction with disdain.

We classified the interactions tagged as "Communication" into primary and secondary themes further refining the nature of the communicative interaction. Primary themes of communication included communication interactions that could be classified as verbal, physical or emotional[1]. Primary themes include verbal, physical and emotional communication were derived and found to be some combination of the primary modes of communica-

[1] Our dataset included only the title or name of the interactions. Coders did not have access to the associated animation, expressions or other media associated with the same and were free to interpret the same. For instance, coders interpreted the interaction, "Threaten," as including the tag *#Physical*.

Table 5.2: Overview of the Communication themes

Primary Themes	Verbal	*e.g. Greet a character*
	Physical	*e.g. Hug a character*
	Emotional	*e.g. Console a friend*
Secondary Themes	Queries	*e.g. Ask someone out*
	Gestures	*e.g. Throw drink in face*
	Mixed Modes	*e.g. Bragging*

tion. These included gestures, queries, and a final category termed as mixed communication. Secondary themes include queries, gestures and mixed modes of communication.

5.1.1 Primary Theme - Verbal

We defined verbal communication as any product of spoken or written language used between two or more social characters to communicate. For instance, we see two examples of interpersonal dinner conversations amongst friends in Fig. 5.1. Approximately 75% of the communication interactions across all projects surveyed were tagged as Verbal. This included interactions such as flirting, asking about one's day, yelling at another character, and so on. A high-frequency code associated with the Verbal tag is #RelSocial#Increase that indicating coders felt Verbal interactions produced an improvement in social relationships.

5.1.2 Primary Theme: Physical

These communications were perceived to be or have an effect on a tangible, sensation (as opposed to verbal, or emotional). For instance, pushing someone in anger, physically flirting with another, or the act of mutiny on a ship. Some of the high frequency codes associated with the Physical tag were *#RelSocial#Increase* indicating that coders felt physical interactions influenced social relationships positively; *#RelRomance#Increase* indicating that the tagged physical interactions were likely to increase romantic relationships; and *#Violence* indicating that physical interactions were also correlated with roughness or cruelty (for instance, punching another character).

5.1.3 Primary Theme: Emotional

Relating to communication of the emotions and feelings of a character to another, or interactions that subsequently affect the emotions of another character. Interactions of the

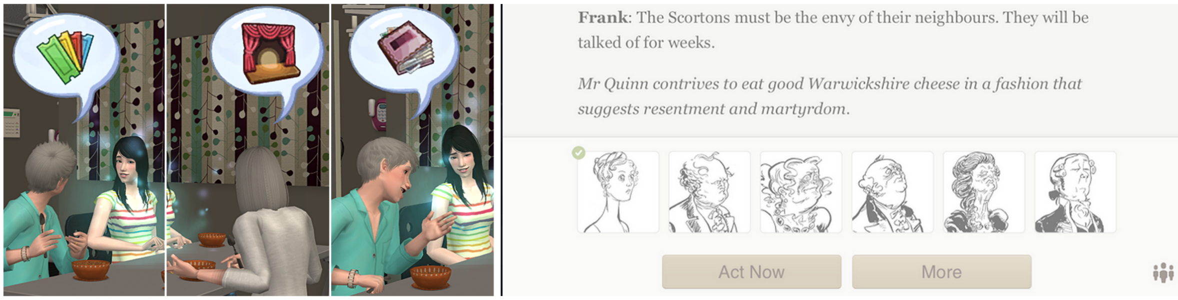


Figure 5.1: Two interactions of dinner conversations. The left interaction plays from *The Sims* (Maxis 2003) where communication occurs primarily through icon images representing topics. The other in *Versu* (Evans and Short 2014), with conversation depicted primarily through text. However, both interactions are examples of Verbal, Bi-directional Knowledge Propagation and will influence the social relationships amongst the participants.

former type, communicating a character's internal emotions could include expressing joy about a job promotion or announcing a pregnancy or taking an angry poop (Electronic Arts 2000). In contrast, interactions that affect another character's emotional state could include apologizing to another character, or appealing to the kindness of another. The emotion was expressed in varying ways across projects using emoticons, a textual representation of joy, and change in animation with expressions, style of walking, or posture. These account for the "facial expression" part of Lisetti (2002)'s taxonomy (Lisetti 2002). A few high-frequency codes associated with the Emotions tag are *#Appreciation*, indicating a character was being appreciative of another, with the effect of improving the social relationship between characters (i.e. *#RelSocial#Increase*) (Electronic Arts 2000; McCoy et al. 2011a; Pynadath and Marsella 2005) and interactions involving specific emotion types such as happiness, being scared, sadness, and romance in decreasing order of frequency.

5.1.4 Secondary Theme: Gestures

Gestures are used primarily in conjunction with physical communication. They are defined as a form of non-verbal communication or non-vocal communication in which visible bodily actions communicate particular messages, either in place of or sometimes in concert with verbal communication. For instance, waving a friend over, greeting someone at the door, or throwing a drink at another. Some high-frequency code associated with the Gesture tag were *#Gifts*, indicating that a common gesture could be seen as the giving of a gift; *#Neutral* or *#Ignore*, indicating a neutral feeling, or a character having no feeling for or against a subject or to intentionally disregard, or refuse to interact with other characters. For instance, making a neutral utterance or comment (McCoy et al. 2011a; Evans and Short 2014), or bidding another goodbye (Electronic Arts 2000).

5.1.5 Secondary Theme: Querying

Queries were used primarily in conjunction with verbal communication while trying to obtain information, and generally were ways for knowledge and information to flow through the simulation. Queries tended to be either verbal in nature, or have a physical component or a gesture associated. For instance, an example of a verbal query could be one character querying another for the location of an object or person (Meehan 1977; Ryan et al. 2015; Nintendo 2001; Electronic Arts 2000), or inquiring about their relationship status (Electronic Arts 2000). An example of a physical query could be characters getting down on one knee to propose marriage (Electronic Arts 2000; Ryan et al. 2015).

5.1.6 Secondary Theme: Mixed Modes

Some communication interactions did not fall neatly into one of the categories listed above. For these interactions, we concluded there were many ways to accomplish the interaction goal, and either projects implemented the same interaction in different ways or did not give enough information for us to classify them into a specific mode. We classify these exceptions as mixed-mode communications, including bragging, embarrassing oneself, reacting to ghosts, and expressing that one is impressed by another character.

5.2 Theme: Flow of Knowledge

Several research groups are interested in simulating the phenomenon of the flow of knowledge through a system. This phenomenon has long been associated as a source of drama and interest between characters and generates a large amount of storytelling (Eger 2020; Azad and Martens 2019). Examples of such narratives in our dataset include social characters forming plans to obtain objects, influencing relationships, forming alliances, or even enacting revenge based on information they learnt throughout their lives.

We tracked the specific changes to the characters' knowledge base and the world to classify the interactions under the sub-themes of Knowledge Creation, Knowledge

Table 5.3: Overview of Flow of Knowledge in our taxonomy. Knowledge was categorized into the creation of knowledge, propagation of knowledge, and termination of knowledge.

Creation	**Propagation**	**Termination**
By the Agent - Invented - External Observation - Introspection or Evaluation *e.g. starting a business*	Type of Propagation - Circulation of information - Using influence of persuasion *e.g. share hobby*	Deterioration or termination of knowledge over time *e.g. forgetting information*
By the System *e.g. news broadcast*	By Direction - Unidirectional propagation - Bidirectional propagation *e.g. debate politics*	
Through social interaction *e.g. eavesdropping*	Veracity of knowledge - Truth - Unintentional misinformation - Wilful lies *e.g. lie about job*	

Propagation, and Knowledge Termination as depicted in the overview in Table 5.3. These categories have been discussed in greater detail in this section.

5.2.1 Knowledge Creation:

We define *Knowledge Creation* as any interactions to do with the addition of new knowledge, either into the simulation or to the social character's knowledge base. On further analysis, we found that these interactions could be further classified based on whether new knowledge was generated or discovered by the NPC, generated by the system, or generated through interactions with other NPCs. We define these categories further below:

- **Generated By the NPC:** Artifacts dealt with the production of knowledge in varying ways. While some chose to instill all known facts into the system at the start of the simulation (Meehan 1977), others chose to allow for the creation of new knowledge or even languages during the simulation (Ryan 2016). We discuss the themes we discovered below.
  - **By Characters:** This includes knowledge generated by the character directly. This could include interactions such as choosing a name for a baby (Azad and Martens 2019), discovering and naming a new island (Ryan 2016), or a skilled product

from an NPC (e.g. a painting by an artist, or a musical composition) (Electronic Arts 2000).

  - **By External Observation:** Social characters can also generate information based on their observations of the environment. An example of this would be an NPC noting the whereabouts of a character (Ryan et al. 2015; Nintendo 2001; Si et al. 2005; Electronic Arts 2000), or an object they desire to obtain (Meehan 1977).
  - **By Introspection or Evaluation**: This includes interactions where knowledge is created through the process of introspection. For instance, determining one's political leanings or ideologies (Azad and Martens 2019), or knowledge unintentionally created by concocting new knowledge through confabulation or transference (Ryan et al. 2015). It could also include a character evaluating another NPC's work, abilities, or skills (Evans and Short 2014; Electronic Arts 2000; Si et al. 2005; Nintendo 2001; Ryan et al. 2015), the interpretation of another character's actions, or even the discovery of new likes or dislikes (be it of items, hobbies, or other NPCs).

- **Generated by the System:** Some knowledge was inserted into the system by the authors or generated by the system itself and then propagated by the social characters. Examples of such knowledge include a random event occurring at a workplace (Electronic Arts 2000), news events found through a newspaper subscription (Azad and Martens 2019), or a new fishing challenge announced by a tradesman (Nintendo 2001)

- **Generated through Social Interactions:** Knowledge was created through the interaction amongst characters. For instance, the mutation, propagation or creation of languages (Ryan 2016), or information learned from eavesdropping on others, or by noting of relationship or affiliations of another character (Nintendo 2001; Electronic Arts 2000; Evans and Short 2014; Azad and Martens 2019; Ryan et al. 2015; Si et al. 2005)

5.2.2 Knowledge Propagation:

Interactions that dealt with the exchange and flow of knowledge through the simulation. Knowledge propagation was further categorized as follows:

- **By Direction:** *Unidirectional* propagation of knowledge occurs for instance when an NPC obtains information from a source. For instance, by directly querying an

NPC (Electronic Arts 2000; Azad and Martens 2019), or eavesdropping on a conversation (Ryan et al. 2015). *Bidirectional* propagation of knowledge could occur if all participating characters shared or exchanged information, for instance, deep conversations (Electronic Arts 2000) sharing the character's likes and dislikes, or sharing of news or personal opinions (Electronic Arts 2000; Azad and Martens 2019).

- **By Veracity:** We found that the information the characters propagated could be further classified based on its veracity. Characters could share the news that was truthful, unintentionally propagate a confabulation, or even tell an outright lie (Meehan 1977; Ryan et al. 2015; Azad and Martens 2019). In addition, characters were also seen to bend the truth when trying to fit in with groups (Azad and Martens 2019), or based on their perceptions of the expectations of the participating dialogists (McCoy et al. 2011a). The decision of the veracity of the knowledge to impart was found to be made based on the decision-making algorithm of the work. This decision varied based on the relationships of the participating characters, their intentions, beliefs, and the trust they shared.

- **Persuasion and Influence:** NPCs were able to persuade, influence, or reinforce their views and those of others around them. This could be seen in interactions such as subscribing to a news source (Azad and Martens 2019), defending an evaluation (Evans and Short 2014), persuading a friend for a favor (Meehan 1977), or imposing one's views or positions into a conversant (Evans and Short 2014).

5.2.3 Knowledge Termination:

Finally, knowledge could also deteriorate or terminate over time. The deterioration of knowledge was found to affect the veracity of the knowledge propagated (Ryan et al. 2015). The artefacts we surveyed modeled deterioration in various ways, from characters unsubscribing from a news source or the propagation of misinformation overriding other knowledge (Azad and Martens 2019) to the literal forgetting or corrupting of knowledge learned over time (Ryan 2016; Ryan et al. 2015).

5.3 Theme: Relationships

Across our surveyed artefacts, we were able to see how social relationships were instrumental to extrinsic, nonsocial ends, or as constraints on the satisfaction of individual desires, driving character decision making, and narrative plot (Si et al. 2005; Ryan et al. 2015; Mateas and Stern 2003; Meehan 1977). In addition, we noted the influences on each relationship or the factors and rules that governed or affected the relationship. To select these factors of influence, we looked at each interaction to determine those that could change either the nature of the relationships held by or the social network of the NPC.

Ho (1998) defined the term relationship as a connection existing between people related to or having dealings with each other, with attributes that are more specific, sharply defined, or lasting. They argued that irregardless of socioeconomic or cultural variations, relationships, and relational contexts affect social behavior (Ho 1998). We feel that this definition captures the features discovered in our taxonomy.

With this section, we first identify and classify the *Relationship Types* present in the surveyed artefacts as depicted in Figure 5.2. Next, we were able to identify the more specific *Relationship Attributes* associated with these relationships. These attributes were seen to be the encoding of social norms and customs associated with the specific relationship. Next, we were able to identify the primary and secondary derivatives of the temporal complexities governing the development, growth, and decay of *Relationship Dynamics*. Finally, the surveyed artefacts were seen to have additional *Relationship Dimensions* representing the individual participants' perceptions of the relationship, and helping us to identify the distinctions between similar relationships. An overview of the attributes, dynamics, and dimensions has been included in Table 5.4 for easy retrieval and understanding.

5.3.1 Relationship Types

We created a tree structure to depict the classes of relationships in our taxonomy and make them more easily understandable. This can be seen in Fig. 5.2. A relationship can be formed on multiple bases, for instance, a character may be a sibling and a colleague to another character if they worked at the same family-owned business. Thus, we define the relationship between any two NPCs to be an overlapping subset of classes. The surveyed projects were found to have some subset of the following relationships:

- **Familial:** Familial relationships consist of direct family connections – for instance, parents, children, spouse/partner – and extended family – for instance, grandparents, in-laws, grandchildren, etc. This could be further differentiated into biological and chosen family. Some codes associated with the forming of new familial relationships are *#Communication*, *#Emotions*, *#Happiness*, and *#Baby*. Example interactions

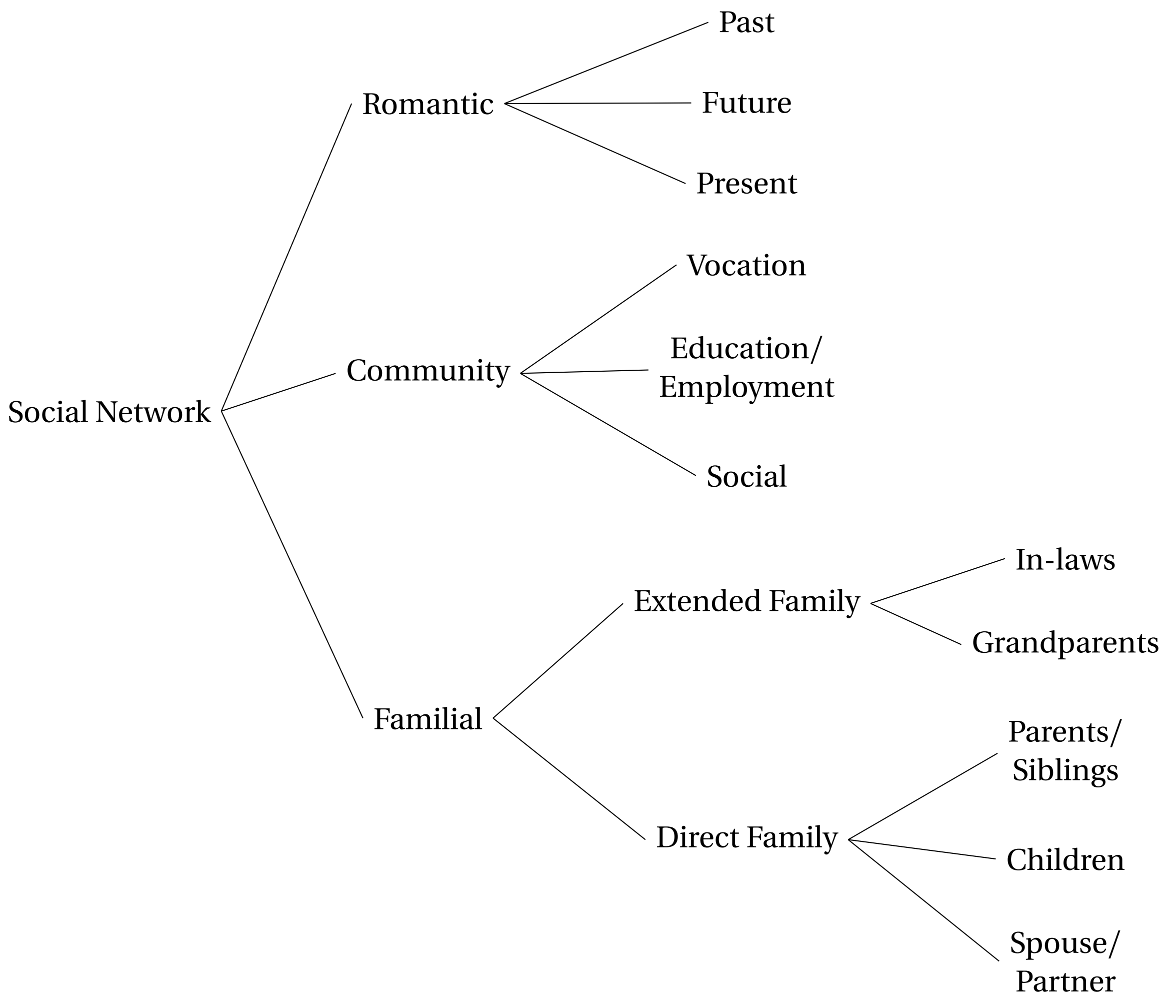


Figure 5.2: Categories of relationship types identified during our review

tagged with these codes include having a baby, getting married, spending time with family, talking to family, asking for help with homework, and adopting a baby.

- **Romantic:** Romantic relationships were handled differently across projects surveyed. Overall they could be divided into an overlapping set of present or ongoing relationships, future ones, for instance, where a spark of romantic interest seemed to be budding or was acted upon, and past relationships. Some high frequency codes associated with the forming of new romantic relationships are *#Physical*, *#Gesture*, *#Romance*, *#Happiness*, *#Emotions*, and *#Verbal*. Example interactions tagged with these codes include kissing, holding hands, asking another character out, and sexually propositioning a character.

- **Community:** Coders tagged all other social relationships under community relationships. This could include social relationships, for instance, those forged with neighbours, or friends; educational or workplace relationships amongst classmates or colleagues; and vocational or affiliation-based relationships, including those amongst NPCs sharing the same hobby, for instance, gym buddies, or with local business owners or community members interacted with. Some high frequency tags associated with the forming of new Community relationships *#Hobby*, *#Trade*, *#Employ*, *#Work*, *#Verbal*, *#School*. Example interactions tagged with these codes include asking about a character's career, visiting a neighbour, establishing a settlement, cheering up a character, giving a gift, and sharing a mutual interest.

With these classes, our taxonomy aims to be generic enough to record simple relationships between a few characters in a smaller simulation and the more complex societal or group connections in larger environments. Based on the project goals, these relationship states may be added, deleted or modified with (Evans and Short 2014) or without the knowledge of the characters involved (Ryan et al. 2015; Azad and Martens 2019; Electronic Arts 2000).

Table 5.4: Overview of the features of Social Relationships in the taxonomy.

Attributes *Encoded social norms, phenomenon, constraints and expectations.*	**Dynamics** *Temporal factors, or dynamics determining the strength of the relationship*	**Dimensions** *Internal differentiating factors and perceptions of participants in a relationship.*
- Acceptability - Exclusivity - Cardinality *(one-one, many-one, one-many, many-many)* - Symmetry - Membership - Volition - Available Behaviours	- Valence - Duration or Permanence - Change in Valence *(Non Recurring, Constant, Accelerated, Unchanged)* - Periodicity	- Trust - Deceptiveness - Competitiveness - Indebted Towards - Power and Domination - Likability - Social Rank - Attractiveness - Compatibility

5.3.2 Relationships Attributes:

Relationships can have various attributes associated. These attributes were seen to be the encoding of social norms and representative of the social phenomenons associated. The discovered attributes are thus constraints on the relationship, and have been listed below:

- **Acceptability:** All the projects incorporated notions of social norms of accepted or unaccepted relationships in their implementations. However, unlike real life, where a social norm can be flouted or disregarded, these norms are programmatic exclusions. Thus, for instance, cinematic or narrative moments with disturbing reveals of incest (Kelly 2019) do not occur. Instead, the projects forbid these interactions from occurring. This leads to scenarios where players are surprised if they cannot flirt with another character, only to realize later they may be distant cousins (Hernandez 2015).

- **Exclusivity:** Whether the relationships and attachments are exclusive, or whether they are general. For instance, one may have only a single biological mother. This is not a relationship that can be broken or exchanged for any other. Alternatively, one may have an aggregative relationship among several colleagues or friends.

- **Cardinality:** Relationships may have a cardinality associated. For instance, the relationship may be a one-to-one relationship (for instance, that between a husband and wife), many-to-many (for instance, group affiliations in a work environment), or one-to-many (for instance, a mother may have several children) in nature. These attributes tell us whether the relationship is considered a unique one, or is a more general one. For instance, in a social system that encourages one-to-one relationships between a husband and wife, characters may have just one married partner at a time. To start another such relationship, a social character would need to end the former.

- **Symmetry:** Symmetry in a relationship is defined by whether both parties in a relationship classify or value it similarly. Relationships can be symmetric; for instance, A's views on B may be considered to be precisely the same as B's views on A. In contrast, for asymmetric relationships, X may regard Y differently. For instance, X may regard Y in two lights, as a friend and a potential lover; however, Y, in turn, may despise X and want nothing to do with them. Such a disparity in relationships can lead to narrative drama and tension, and is one of the story patterns that is recognized and showcased by story-sifters. In this case, any romantic advances of X may be rebuffed and scorned.

- **Membership:** Relationships can also form between social characters that share the same membership, affiliation or interest as another. For instance, the relationship between two individuals with the same political affiliation, or belonging to the same cultural group or ethnicity. These relationships were found to give members a feeling of security or being accepted, as well as have respect or loyalty associated.

- **Volition:** Relationships may be voluntary, for instance, the bonds of friendship formed amongst neighbours, or involuntary, for instance, the relationship amongst siblings, or that in an unwanted, or arranged marriage.

- **Available Behaviours:** Once a character begins a relationship with another (whether symmetric or asymmetric), certain behaviours may become available to them. For instance, allowing a character in a budding romantic relationship to flirt with their love interest (Electronic Arts 2000; Evans and Short 2014) or to give a gift to a friend (Electronic Arts 2000; Nintendo 2001).

These constraints can affect the character's behaviour with one another, encouraging or dissuading certain sets of interactions that could be tagged with these constraints. For instance, a married character may not flirt with another character, or one fancying themselves in love may propose to their loved one. A situation where these encoded social norms or constraints are broken would lead to narrative tension. For instance, generating a narrative with one character cheating on their spouse.

5.3.3 Relationship Dynamics:

Every relationship has temporal factors or dynamics associated with it that determine the current strength of the relationship. We divided these discovered factors into two categories: the primary derivatives and temporal factors that affect the duration, charge or periodicity of the relationship, and the secondary or second derivative of these factors. Relationship dynamics can be mixed and matched with any specific relationship between two characters being studied and help to track the ebb and flow of their relationship over time. These have been described in further detail below.

- **Valence:** Artifacts surveyed encoded the strength of the relationship reflecting it's permanence or eventual termination by associating with it a valence. A valence is denoted by an intrinsic positive or negative quality in the relationship. The valence of a relationship can be asymmetric, or one participant in the relationship may judge the valence or strength of the relationship differently than the other.

- **Duration or Permanence:** Relationships were found to have varying degrees of permanence. This could range from one-time or situational encounters and consultations, for instance hiring an architect (Ryan et al. 2015), or those of temporary tutelage or apprenticeship (Azad and Martens 2019), to more permanent ones, such as those between close friends or family (Electronic Arts 2000).

- **Change in Valence:** Artifacts were able to programmatically encode the social phenomenon of the change in the intensity relationship between characters. They achieved this by including a numeric quantity that charged or decayed the valence of the relationship. The change in valence is asymmetric and based on the participant's perceptions of the relationship. The degree of difference allows for interesting narrative scenarios. For instance, more outgoing characters vs introverted characters could have differing social needs. Thus, a sudden charge in a relationship for an extrovert attending a party might correlate to a decay in the same for an introvert. Some high frequency codes associated with the charge of a relationship are *#Gift*, *#Errand*, *#Appreciate*, and *#Entertain*. In contrast, coders attributed the decay in a relationship to interactions coded with *#Fight*, *#Violence*, or *#Mean*. To encode this change in valence, we list the secondary derivatives of relationship valence dynamics:

  - ***Non-Recurring :*** A relationship could undergo a one-time enhancement or decrease in the charge based on a single interaction. For instance, being the recipient of a rude gesture may involve a one time decrease in the relationship's valence.

  - ***Constant:*** A steady decay or charge in a relationship based on periodic interactions amongst characters.

  - ***Accelerated:*** Acceleration in the charge of decay caused by a specific event or circumstance that has occurred. For instance, going out for drinks with your colleague could charge the valence acceleration for a co-worker relationship.

  - ***Unchanged:*** No change in the relationship amongst characters.

- **Periodicity of relationship:** Defining whether a relationship once ended can restart, or how past relationships can influence the new. For instance, relationships involving ex-lovers that get back together, remarrying an old spouse, or friends reuniting after several decades.

5.3.4 Relationship Dimensions:

The surveyed artefacts were seen to have additional dimensions representing the individual participant's perceptions of the relationship. These further dimensions help us to identify the distinctions between similar relationships. Evans and Short (2014) describe these as role evaluations that are taken into account. These factors were found to vary with projects based on the social phenomenon being evaluated, the social environment in which the characters are situated or the individual attributes or personalities encoded in the work. For instance, in *The Sims*, characters may have a proclivity to admire characters with blue eyes (Electronic Arts 2000). Similarly, in Versu, the Bingley sisters evaluate Elizabeth on her style, moneyed relations, beauty, lack of skill on the pianoforte (Evans and Short 2014) while judging her as a friend, and Elizabeth evaluates Darcy as a romantic suitor based on how well-bred he is, his attractiveness, and his rude nature vs the affable nature of Mr. Bingley.

Our taxonomy identifies some of the common personal dimensions evaluating these relationships. Each dimension can be seen as the character's evaluation of the other. Projects may also include character mental models, with characters reasoning about the reciprocity, intentions and evaluations of their social network (McCoy et al. 2011a). For the purpose of this research, we limit our discussion to the factors themselves and not the associated mental models available. These dimensions include but are not limited to:

- ***Trust:*** How trustworthy is the other?
- ***Deceptiveness:*** How deceptive is the other?
- ***Competitiveness:*** How competitive is the other?
- ***Indebted towards:*** Is this character indebted towards the other?
- ***Power and Domination:*** What is the power dynamic between the characters? Does one dominate the other?
- ***Likability:*** Is this character likable?
- ***Social Rank:*** What is the social rank or status of the other ?

Table 5.5: Overview of the modeling of Emotion in our taxonomy

Type	Principle of Contingency	Principle of Inertia vs Principle of Regulation	Principle of Interaction
Happy, Angry, Sad, Worried, etc	Emotions are responses to extrinsic events called moodlets.	Emotions have inertia and must be regulated to maximize utility.	How the components of emotions interact with, augment and blunt one another
	- Antecedent Cause - Emotion Type - Valence	- Inertial Duration (e.g. 10 mins) - Regulated Effects (e.g. apologize to reduce mortification)	- Composite - Exclusive

- ***Attractiveness:*** How attractive does this character find the other?
- ***Compatibility:*** How compatible are the characters?

5.4 Theme: Emotions

It is critical while designing agents that interact with other agents or humans that we consider the emotions involved. Emotions play a powerful role in social influence and interactions (Gratch and Marsella 2005; Yannakakis and Paiva 2014). Emotions can change the tone and intensity of social interactions (Evans and Short 2014). Similarly, the display and cognizance of emotions help improve social utility or happiness by minimizing social or cultural conflicts. The exercise of accurately modeling emotion can often spur the development of new mechanisms that may be of general use to agent systems. For instance, Mao's effort to model anger led to a general mechanism of social credit assignment and a model of social coercion (Gratch and Marsella 2005; Mao and Gratch 2004). This improves social utility by minimizing cultural conflicts (Lisetti 2002; Yannakakis and Paiva 2014; Pynadath and Marsella 2005).

We used a model-free approach to classify expressions of emotions and the factors associated in our dataset. During our analysis, we realized that the factors and principles we identified in our dataset are those commonly associated with the field of Emotional dynamics (Kuppens and Verduyn 2017). We re-structured our discovered factors by these principles, and note how our surveyed artefacts incorporated them into development below.

For the purpose of illustration, we will use the running example of a character responding to a sudden death (or murder) of a loved one to describe the factors below (Evans and Short 2014; Ryan et al. 2015; Hernandez 2015).

5.4.1 Identified Emotion Types:

Several computational models of human emotions have been developed in the past (Marsella et al. 2010; Ekman 1992). For the scope of this work, we looked at the emotional states most prevalent in our surveyed projects. We tagged our list of interactions with the emotions resulting from or associated with the interaction.

From our coded interactions, we were able to shortlist the emotions associated commonly with the interactions to the following: *Happiness, Anger, Sadness, Worry, Envy, Appreciative, Disparaging, Bragging, Scared, Neutral, Ignored, and Rebelliousness.*

In the case of our example, a character may be dealing with sadness or grief caused by the death of a loved one.

5.4.2 Principle of Contingency:

Characters did not randomly move into emotional states, but instead in each case reacted to a change in external, or internal social environments. The one exception to this cause-effect emotional state is that in *Animal Crossing*, where an NPC may be *Cranky*, *Irritable*, *Perky*, and so on by default as their personality (Nintendo 2001).

Prior research defines this observation as the *Principle of Contingency* of Emotions. Simply stated, emotions consist of responses to things extrinsic to them. They can be contingent on internal or external events, often social in nature (Kuppens and Verduyn 2017).

In *The Sims*, the Principle of Contingency is programmatically achieved by a creation of a *Moodlet* (Maxis 2003) for any external event that causes an emotion or mood-altering circumstance. *Moodlets* detail their cause, and the effect experienced by the character. For the purpose of our taxonomy, we adopt this term. An interaction may produce multiple *moodlets*, each affecting a separate emotion of the character. We consolidate the representation of these *moodlets* across the surveyed systems[2]. We define a *moodlet* to consist of a

[2] A version of Prom Week's CiF was paired with Fear Not! Affective Mind Architecture (FAtiMA), an agent architecture extension of the BDI (Belief, Desires, Intension) and OCC model, that treats emotions as valenced evaluations of the world integrated with coping mechanics (Lim et al. 2008). However, this version of the system was not incorporated in our taxonomy

single causal factor, modifying valence, and type of emotion associated with any emotion modifying extrinsic event.

- **Antecedent Event or Focality:** Specific internal or external event or interaction causing the emotional effect (Russell 2003; Lisetti 2002). Further, it is necessary to note that a character may not correctly identify or attribute the antecedent event to be the cause of the emotion. In our example, the traumatic event of experiencing or witnessing the death of a loved one would be the external causal factor.

- **Emotion Type:** We found that coders tagged interactions with multiple emotions, indicating the same interaction could have multiple effects on a character's mood. Thus, an interaction could add more than one *moodlet* to the character's set of emotions. For instance, witnessing death could add an Anger *moodlet* (blaming a suspect) and a Sadness *moodlet* (attributing grief to the felt loss) to a character (Evans and Short 2014; Eger 2020).

- **Valence:** Every *moodlet* is associated with an effect or valence by which it adds or alters the current emotion experienced by the character. In the case of our example, witnessing death could increase sadness in a character by a unit of, say, +20. In this case, valence corresponds to the horizontal pleasure-displeasure scale in Russell (2003)'s theory (Russell 2003). *The Sims* (Maxis 2003), further relates these valences to ranges comparable to a mapping of the intensity (Lisetti 2002) of the associated emotion. Every emotion has a range and a superlative for the same associated with it. If the emotion is not dealt with, a character may die of emotional exhaustion. For instance, a sim could be embarrassed, very embarrassed, mortified and then die of mortification. This mechanic has lead to very dark, but fascinating and chronicled user-generated narratives (Hernandez 2015).

5.4.3 Principle of Inertia vs Principle of Regulation

Emotional states display an intrinsic resistance to change, even in the presence of forces that motivate change, causing them to display a general inertia, or tendency to carry over from one moment to the next (Kuppens and Verduyn 2017). Additionally, emotions are continually regulated to maximize fit with the current desired state. These two factors act together influencing the duration and valence of the emotions associated.

To represent this phenomenon, artefacts were found to have a duration associated with their *moodlets*. Additionally, while in the emotional state, certain reactions (Evans and

Short 2014), or interactions were found to be available to the character that are specific to that emotional state that could influence or regulate the emotion.

- **Inertial Duration:** Similar to relationships, *moodlets* have a duration associated. For our example, the sadness could persist for up to a week (Evans and Short 2014; Ryan et al. 2015). During this time, the entire effect of the *moodlet* may be felt by the character (Electronic Arts 2000), or the effect could decrease over time. Our survey did not unearth sufficient proof to recommend a method to recommend how to handle *moodlet* decay.

- **Regulated Effects:** Some agent behaviours or interactions are only available in specific emotional states. These behaviours are seen to regulate the emotional states, and affect a change in the intensity of the experienced emotion or help resolve the situation. Programmatically, this was implemented by some interactions have emotional preconditions associated with them. This was seen to manifest across the artefacts explored. For instance, in *The Sims*, Taking an Angry Poop (available when angry) reduces the intensity of anger. In another project, a character may choose to dissolve a relationship or leave a job that made them unhappy regulating the effects of drawn-out depression (Ryan et al. 2015, 2016a; Electronic Arts 2000; Evans and Short 2014). Uniquely, *The Sims* allow characters to die due to being in an extreme or unregulated emotional state for a prolonged time. For instance, one can die of mortification.

5.4.4 Principle of Interaction:

We found that interactions in our dataset had overlapping emotional states associated with causal events. Multiple emotions were found to react together to produce an overall emotion or mood for the character. Further, characters were able to be in multiple emotional states at a time.

Kuppens and Verduyn (2017) describe how the components of emotions (physiological, experiential, and behavioural), or the emergent emotional states as they are experienced, continuously interact with, augment and blunt one another, creating a system that displays evolving patterns of synchrony and networks of interacting elements. We categorized them as follows:

- **Composite:** In *The Sims* (Maxis 2003) universe, characters can experience multiple simultaneous emotional states. The emotions and their associated *moodlets* are not exclusive, so, for instance, it is possible to be in a positive mood from "Beautiful surroundings" and a negative one from "Vile surroundings" from another object at the same time.

- **Exclusive:** A simple model of emotion where one emotional state may persist for a short period; however, it will be overridden by another. Thus, a social character may not experience more than one emotion at the same time. For instance, in Versu (Evans and Short 2014) while characters remember their earlier emotional states, the system can *shift* them to a new emotion.

Interestingly, both Versu (Evans and Short 2014), categorized as having an exclusive emotional model, and *The Sims* (Maxis 2003), categorized as a composite model, discuss how certain events and objects can *enhance* the intensity of already existing character emotions (Nardone 2014; Evans and Short 2014). For instance, looking at a photograph reminder of a lost one may enhance sadness, or gussying up may make one feel more optimistic about a day. However, they both point out that while these events enhance emotions' valence, they cannot change existing emotions completely. For instance, a character may not laugh if they are already in a bad mood.

CHAPTER

6

CASE STUDY: ANTHOLOGY

We learnt from the problems we faced during the design and development of Lyra. Having already begun surveying other existing social simulations and designing the Little Computer People Taxonomy, we began to work on our second case study and social simulation, Anthology.

Social simulation research seeks to understand the dynamics of complex human behavior by simulating populations of individual decision-makers as multi-agent systems. Despite the success of using social simulation as a core aspect of gameplay, there was a seeming lack of publicly available tools for helping game developers create these types of experiences (Johnson-Bey et al. 2022a).

With Anthology our goal was to create a extensible, expressive, and reusable social simulation toolkit that could be used to examine underlying decision-making or computational social science theories. However, prior work in games and entertainment fail to account for social interactions, geography, and social relationships in a manner that allows researchers to easily reuse their frameworks and model social characters. Instead, as we described in Chapter 3, existing social simulations were too tightly enmeshed and interwoven with one another to be successfully integrated or reused for a new application. In the past, that meant that we had to reinvent the wheel, and design Lyra from scratch. Our primary goal, was that Anthology users would not face this issue.

To achieve our goals we undertook a design approach using a human-centered software development perspective. We wanted to design a usable, open-source implementation, alongside clear documentation, examples, and instructional materials for running one's simulation, to enable better reproducibility and reuse, obviating the need for other research groups to reinvent the wheel. This chapter describes the participatory research design process we undertook to design Anthology's features, the state of the system, the results of an expert evaluation of Anthology's usability and usefulness for social simulation in practice, and the future directions for this project.

Anthology comprises a motive-based agent decision making algorithm; a knowledge representation system for relationships; a flexible specification language for precondition-effect-style actions; a user interface to inspect and interact with the simulation as it runs in real-time; and an extensive user documentation and reference manual.

6.1 Introduction

Computers are increasingly being used to simulate and analyze complex social phenomena to provide models that can help predict and explain human behavior. Simulating human societies allows more rapid and less disruptive social experimentation: one can change the parameters of a simulation and observe its effects without requiring decades-long longitudinal studies or causing any harm.

However, social systems are complex and contextual in ways that are not captured by current techniques. For instance, most social network simulations (McCoy et al. 2012) do not account for agent decision-making outside of the social network, including day-to-day activities such as going to work, discussing one's life with family members and friends, and leisure time activities, all of which can influence human social behavior (Azad and Martens 2019). Conversely, simulations of strictly geographical phenomena (such as transportation traffic or land-use change) do not account for the social relationships among humans, including family, work, school, and community relationships, that are inextricably tied to how people decide where to live and how to transport themselves between locations.

This paper presents a system design and ongoing research agenda to integrate social AI techniques from social simulation games (McCoy et al. 2012; Samuel et al. 2015; Evans and Short 2014) into an extensible framework for social system modeling that incorporates both geographical and relational factors. Our system, Anthology [1], comprises a motive-based agent decision-making algorithm; a knowledge representation system for relationships between agents and other world entities; a flexible specification language for precondition-effect-style actions that can both depend on and modify said motives and relationships; and a user interface to inspect and interact with the simulation as it runs in real-time.

Our goal is to provide a usable, open-source implementation alongside clear documentation, examples, and instructional materials for running one's simulation to enable better reproducibility and reuse, obviating the need for other research groups to reinvent the wheel. For this reason, a core contribution of this work is the presence of a thoughtfully designed specification format for configuring and extending Anthology's basic functions. Users provide a world specification file in a standard format (Javascript Object Notation, or JSON) that defines agents, locations, and actions (akin to a PDDL file for planning systems). Our base simulation includes a set of motivations or needs that drive the character's decision-making algorithm. However, these can also be modified or added to if desired.

This paper describes our participatory research design process for designing Anthology's features, the state of the system, the results of an expert evaluation of Anthology's usability and usefulness for social simulation in practice, and the future directions for this project. In the long term, we aim for this project to enable reuse and reproducibility for social simulation research projects within and outside of our group and to allow social simulation researchers to model and reason about the complex dynamics of human social behavior.

6.2 Related Work

Anthology is inspired by several previous projects that aim to simulate social behavior in virtual agents and non-player characters.

Prior work includes *Comme il Faut* (CiF) (McCoy et al. 2010a) and its successor project *Ensemble* (Samuel et al. 2015). CiF boasts the first implementation of a "social physics engine," simulating narrative scenarios using social rules. Anthology incorporates similar rules in the form of requirements for actions which describe social preconditions that must be met for the actions to be available to the agents. However, CiF was primarily employed for a single game development project, Prom Week (McCoy et al. 2012), has no formal definition outside of its implementation, and lacks comprehensive documentation for use

[1] Code repository, API documentation, and examples can be found: https://github.com/SashaZd/Anthology-Social-Simulation-Framework.git

in external projects, creating barriers to reproducibility and reuse, whereas those features are key priorities for Anthology.

The interactive narrative authoring tool *Versu* (Evans and Short 2014) has agents that use utility-based decision making, considering both social norms and their own goals. In Anthology, agents also use utility-based decision making, seeking the actions that maximize their motives. Unlike Versu, which does not have a built-in locative component, Anthology agents also consider geographical context, constraining actions to specific locations, considering agents present, and travel time when computing utility.

More recently, attention has turned towards the need for authoring tools and languages, such as *Kismet* (Summerville and Samuel 2020), for social simulation. Kismet is billed as a lightweight social simulation specification language for facilitating the creation of small-scale scenarios, such as table-top role-playing games. Kismet was designed to be accessible to lay people (e.g., non-experts in social simulation), yet expressive enough to cover a large range of possible scenarios. Likewise, Anthology users can edit a single JSON file, which can also be written in any text editor. Anthology's goals appear to diverge from Kismet's mostly in audience and intended use: we want to offer a flexible range of utility-calculating algorithms for agent decision-making (whereas Kismet currently supports only a simple "proclivity" model that determines the locations that agents travel to) and support realistic simulations of human populations at the scale of neighborhoods or cities. However, since both projects are at an early stage, there may still be fruitful paths to explore to integrate or better differentiate these efforts.

6.3 Research Goals

Anthology is intended to enable researchers, game developers, and social scientists to model and reason about virtual agent behaviors. Our primary design goals are usability and expressivity, defined as follows.

6.3.1 Usability:

Users should be able to create a working simulation, from scratch, with 2-5 agents and 5-10 actions. It should take less than an hour, and they will not require supervised training from the research team nor need to read any source code.

6.3.2 Expressivity:

Users should be able to conceptualize multiple different use cases for the tool (once they understand its scope) across any genre of simulation, and successfully implement these scenarios.

These goals are not new: agent modeling systems, language specifications, and frameworks for social simulation have been generated in previous work (McCoy et al. 2010a; Evans and Short 2014; Summerville and Samuel 2020). However, in practice, these systems do not see use outside of their original research team. Consequently, framework structure, and model designs are often chosen ad hoc, and the focus is often on how to represent agents without sufficient emphasis on analyzing and validating the applicability of models to real problems (Grimm and Railsback 2005).

A recent survey showed that existing social simulation systems tend to invent new project-specific terminology and do not build on existing conceptual frameworks for social systems, despite implementing very similar sets of social concepts (Azad and Martens 2021). This disparity among approaches impairs the reuse of these systems, with users opting to reinvent the wheel, building their simulations from scratch to test a single decision-making algorithm (Marsella et al. 2004; Azad and Martens 2018). Two key elements of our approach support these goals: *participatory design research methodology* and our *motive-based decision making* approach. These have been detailed below.

6.4 Participatory Design Research Methodology

To ensure that Anthology meets the needs of our intended user base, we adopted a participatory design research methodology and development process, involving prospective users in a co-design and co-production process (Spinuzzi 2005; P. Carvalho et al. 2021; Mirel 1998). That is, rather than conduct human-subjects studies to evaluate the framework only at the end of development, we involve users in its design from its inception, allowing their goals and needs to shape Anthology's architecture and feature set.

Participatory design is iterative: we alternate phases of development and (formative) evaluation with prospective users. For each of our participatory phases, we recruited and interviewed experts familiar with social simulation. We performed an expert evaluation on testable versions of our prototypes after each development phase. We observed how experts used Anthology intending to understand their existing thought processes and workflows,

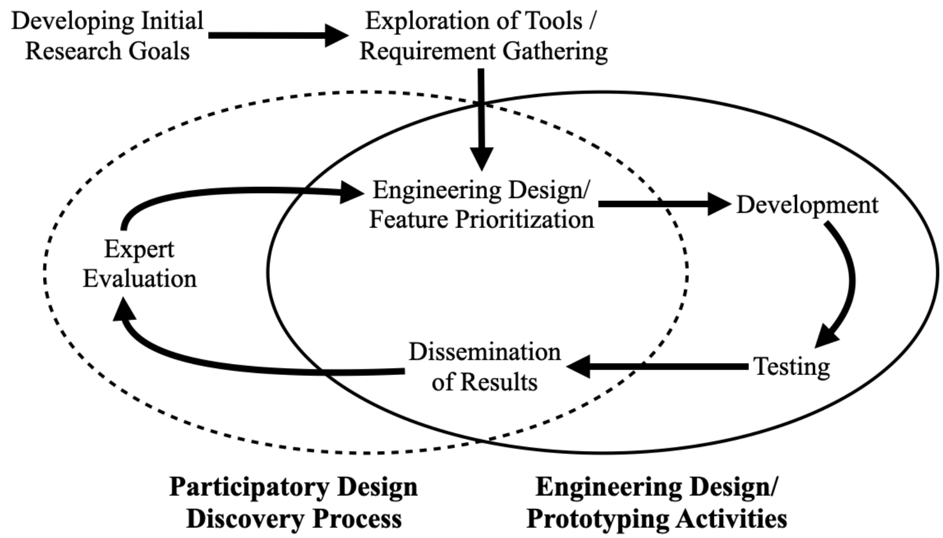


Figure 6.1: Participatory Design Methodology Overview

as well as their *knowledge by doing*, the tacit ways in which they demonstrate knowledge by performing activities (Spinuzzi 2005) to meet their narrative and simulation goals. We noted when and why Anthology frustrated them or helped them accomplish goals. We used this feedback to iterate on each prototype and build the next one, which would again be followed by user feedback, for the entire lifetime of the project. This approach is illustrated visually in Figure 6.1. This process is still ongoing, but we have reached a stable enough point with our prototype that we believe it will serve the broader research community to share it at this stage.

We began by developing initial research goals, identifying the user base, and defining usability and flexibility goals for Anthology (as stated above). Next, we performed a literature review and gathered requirements based on a recent survey of social simulation and agent-based modeling tools (Azad and Martens 2021) During this process, we consulted with a paid expert on agent-based modeling and computational social science who is familiar with building frameworks for both researchers and industry practitioners who want to use social simulation tools to analyze real-world human behavior. These steps allowed us to define the minimal viable product (MVP) features of Anthology required to be developed before any evaluation or co-design phases could begin. We generated and prioritized a list of features to serve as a roadmap for a sequence of Anthology prototypes. Once the MVP was ready, we created comprehensive documentation to support self-directed learning and the use of Anthology.

In our next iteration, we recruited two additional experts in games and social simulation research to begin our expert evaluation Mirel (1998) where we observed experts exploring our tool. In a synchronous study session with each expert, we provided a brief tutorial and a pointer to our web-based documentation. They were asked to perform a series of tasks with the tool, aimed at evaluating the usability and flexibility of Anthology. We started with instructions to complete a scenario we provided and gradually progressed to more open-ended prompts, concluding with a think-aloud task for the expert to imagine any social simulation scenario and attempt to implement it in Anthology, commenting on whether and where Anthology enables or obstructs them from achieving their self-defined goal. This study design allows us to observe experts adapting the software to their imagination, to better understand their expressive goals and the expressive range of our system, and to refine our requirements specification and roadmap for Anthology. It also enables us to benefit from participants' domain expertise, communicated tacitly through the expert's interactions with the tool and resulting scenario examples produced during the session (as described by Mirel (1998)).

We took into account the feedback we collected, and *iterated on the design* of the tool following the software engineering cycle (i.e. design, development, and testing), followed by an updating of our documentation for a detailed *dissemination of results*. Once this process was complete, we iterated a 3rd time, recruiting two new experts (from social simulation research and an industry practitioner) and repeating the process above.

Using the Participatory Research Design methods described above allows us to iteratively design and develop the emerging design. User feedback phases elicit evaluative feedback on the current state of the framework as well as co-interpretation of the design and research, helping us to envision and shape the system alongside participants who will eventually form a part of our user base for the tool (Spinuzzi 2005; Mirel 1998). We have found that this perspective, allows us to leverage the intuitive and tacit knowledge of the experts in the domain, understand their priorities and needs, and develop a tool that is an extension of this knowledge.

We believe that the Participatory Research Design methodology we followed can foster a strong community of researchers, engage our user-base as co-designers, and improve the reuse and reproduction of research. We place our user base at the center of our design process, working with them to create an authoring tool that can capture their tacit domain knowledge, as well as address their goals. We believe this methodology has contributed to the project making sustainable progress, and that in the long term it will improve the practicability and long-term reach of the project.

6.5 Approach

We believe our decision to use Motive-Based Decision Making as our approach supports our design goals of usability and expressivity.

Code Listing 1 `Default Decision-Making Algorithm`

```
function get_next_action(agent):
  best_act = wait;
  best_u = 0;
  for each action of world:
    find locations where action is possible
    compute travel times for each location
    get nearest such location l
    u = utility(agent, action)
    u /= action.time + travel_time(l)
    if (u > best_u || (u == best_u && withProb(0.5))):
       best_act = action
       best_u = u
    agent.action_queue.push(best_act)
```

On the AI architecture side, we decided to base Anthology's approach to agent decision-making on one of the simplest and most widely-familiar ideas in-game AI: motive-based utility. This idea is familiar to many people because of the popularity of the Sims franchise (Maxis 2003). In the Sims, agents display six *motives*—hunger, hygiene, bladder, energy, social, and fun—that must all be maintained at an adequate level to keep a Sim happy. If the player is not actively controlling a Sim, an automated algorithm will kick in to select an action based on these motives and their depletion levels. each action in the world (e.g. watching television or cooking a meal) can fill and/or deplete any subset of these motives. [2]

Code Listing 2 `Motive-Sum Utility Calculation`

```
function utility(agent, action):
  let u = 0
  foreach effect of action:
  u += clamp(agent.motive[effect.motive] + effect.delta,
     MOTIVE_MIN, MOTIVE_MAX)
  return u;
```

[2] Our terminology and formalization of this idea comes from Millington and Funge's textbook *Artificial Intelligence for Games*, 2nd. ed. (Millington and Funge 2018).

Alternatively, a simple utility calculation could be:

Code Listing 3 Simple Utility Calculation

```
\begin{gather*}
utility(action, agent) = \\
\sum_{e\in action.{effects}} clamp(agent.motive[e.m] + e.{\delta})
\end{gather*}
```

As Millington and Funge note, motive-based decision-making algorithms fall under a more general class of approaches to goal-oriented agent behavior (Millington and Funge 2018). Starting with a naive approach that evaluates an action based only on its immediate net effects, one can gradually dial up complexity to handle more complex situations, up to and including general-purpose planning algorithms that generate multi-step action sequences. This unification of approaches gives us a nice way to structure the system modularly, leaving utility calculation (and decision-making more generally) open to any decision-making algorithm that calculates numeric utility for an agent and an action. In our initial prototype, we use a simple utility-sum-based to select actions as can be seen in Code Listing 2 (or a simplified version in Code Listing 3). For each action that is possible for an agent, we consider the sum of its effects on all motives and divide it by a metric to account for the travel time required for to reach the nearest location where they would be able to perform that action. See Code Listing 1 for a more formal description.

6.6 System Description

At a high level, Anthology follows a Model-View-Controller software design pattern and architecture. Anthology is currently able to simulate virtual worlds that can be described using *models* of agents, actions, and locations. In Figure 6.2 we depict how the interactions between the agent, action, and location *controllers* are mediated by a central execution engine. The models are populated by the user by the input of a single JSON file. The input from the JSON file sets the initial state of the system and begins the simulation. The user can *view* the output of the system on our web interface, which depicts a world map, the locations, agents (with their current motivations), and any upcoming events in the system. The javascript console can be pulled up for a detailed log of historical events in the simulation.

We adopt a running example we have devised called the College Roommates Scenario

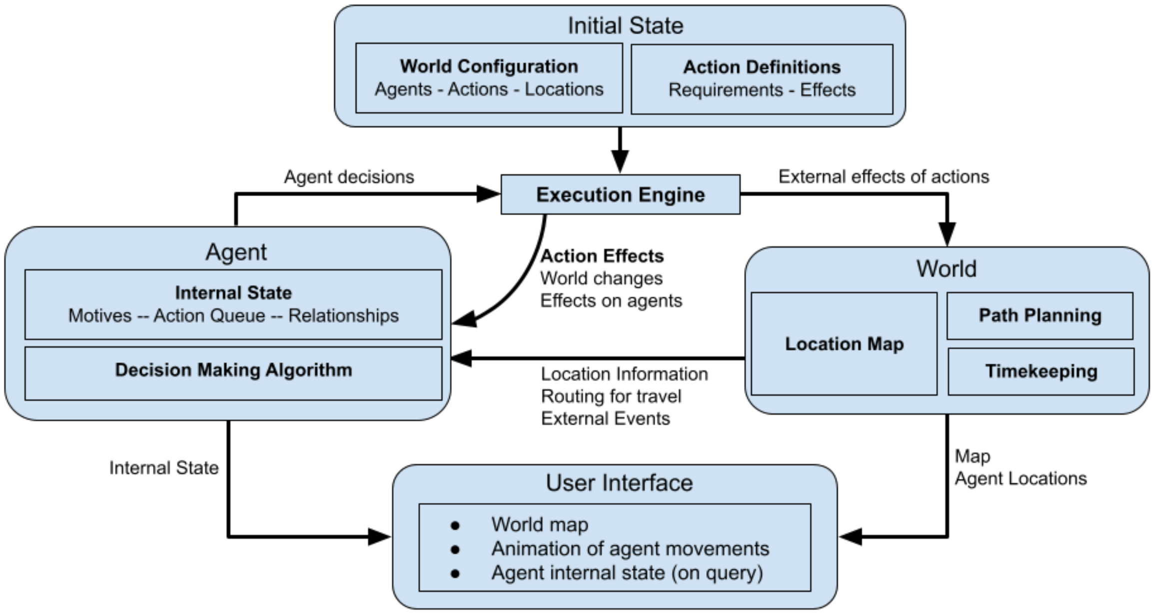


Figure 6.2: Anthology's system architecture.

to explain each component of the system. In this scenario, *Norma*, a math student, and *Quentin*, a physics student, live in the same dorm. The campus on which they reside includes the dormitory, lecture halls, a computer lab, a dining hall, and a greenway (outdoor path). Our modeling goal is to simulate and observe how college students might balance different life activities, such as attending class, doing homework, eating, relaxing, and socializing, by providing simple models of how these actions affect them.

6.6.1 Agents

Agents represent the virtual characters in Anthology simulations. The Agent type includes a unique name, a set of (mutable) motive values, and (mutable) relationships with other agents.

Motives represent the needs of the agents. By default, we supply 5 motive types in the system: Physical, Emotional, Social, Financial, and Accomplishment, based loosely on Maslow's classic theory of human motivation (Maslow 1943). We chose these motives to correspond to real-world human motives as follows:

- *Physical*: The need for an agent to maintain their body through actions such as eating, sleeping, and exercising.

Code Listing 4 Agent Example: `Norma, a math student`

```
{   name: "Norma",
      motive: {
          accomplishment: 2,
          social: 2,
          physical: 4,
          emotional: 3,
          financial: 5 },
    relationships: [
        { type: "friend",
          with: "Quentin",
          valence: 3 },
        { type: "student-of",
          with: "MathProf",
          valence:1 }],
    currentLocation: {
        xPos: 0,
        yPos: 0 },
    occupiedCounter: 0,
    currentAction: "wait_action",
    destination: null
}
```

- *Emotional*: The need for leisure time, play, and mental rest, addressed by actions relating to rest or recreation.
- *Social*: The need to interact positively with other agents, met by actions that involve multiple agents.
- *Financial*: The need for financial stability; addressed by working.
- *Accomplishment*: The need to achieve something, addressed by having hobbies or earning rewards.

Each motive is represented by a number on a scale from 1 to 5, where lower numbers indicate a lower level of satisfaction. Motives change over time, both in response to the actions an agent takes and as a product of motive decay on a fixed interval.

Agents move the simulation forward by undertaking actions. Each turn, they either make progress towards the next action in their queue, or they choose a new action to add to their queue. To facilitate this functionality, agents also maintain state information to aid in the execution of actions: their current location on the map, their current action, and how much longer they must perform it (the “occupied” counter). However, this information is maintained by the execution engine and does not need to be specified by the scenario author. See Listing 4 for an example of how we could model our college student Norma.

6.6.2 Locations

Locations represent geographic points of interest on the Anthology World Map. Anthology uses the locations listed in the JSON file to setup the world, check whether actions can be performed at a specific location, and track agents moving around the map.

Code Listing 5 Location Example: `Dining Hall Location`

```
{
  name: "Dining Hall",
  xPos: 5, yPos: 5,
  tags: ["food"]
}
```

Locations can be added by adding geographic coordinates, an optional name, and an optional list of associated tags to the JSON. The name functions largely the same as it does for the agent and action types. The list of tags represents what actions are possible at a given location. These tags are fully user-defined and are compared against location requirements when agents consider which actions to take. For instance, Norma would evaluate the Dining Hall location (with the “food” tag) as represented in Figure 5 and depicted on the top right corner of the map in Figure 6.3 as a candidate location for dinner.

6.6.3 Actions

In Anthology, actions are the main mechanism by which the agents advance the simulation and change their motive values. While there can be different types of actions within the simulation, all actions include a unique name, an associated time that it occupies, and a set of requirements.

Just as with agents, the name is an identifier used to refer to the action within the simulation and to briefly describe what it represents. All actions have an associated time which represents how many turns (or minutes of simulation time) the action takes to be completed (if not interrupted). Agents only choose a new action if they are not already executing an action. An agent may hold a list of several actions at any given time forming a queue of actions to take where the front of the queue is the action currently being performed. This can be seen in Figure 6.3 where Norma's action queue includes traveling to a restaurant to join a friend who is currently eating their lunch there. In general, there are three types of requirements actions can have:

Code Listing 6 Action example: `Attending Class`

```
{    name: "attend_class",
     requirements: [
       { reqType: "location",
         hasAllOf:["classroom"]},
       { reqType: "people",
         relationshipsPresent: ["student-of"],
       minNumPeople: 2 }],
     effects: [
       { motive: "accomplishment",
         delta: 1 }],
     time_min: 75
}
```

- *People*: This requirement enforces which agents are present or absent for an action to be performed. Authors can specify the minimum and the maximum number of agents attending, list out specific agents that must be present or absent, or require the presence or absence of an agent by specifying a relationship type.
- *Locations*: Location requirements determine the eligibility of a location for an action to be performed. Authors can specify a set of tags that must all be present in the location's tag list, a set that one or more of must be present, and a set that all of must be absent from the tag list.
- *Motives*: Motive requirements allow authors to compare agents' motive values to threshold values to determine whether or not an agent is capable of taking the action.

The action, `attend_class`, shown in Figure 6 depicts an agent, Norma, attending a lecture. In this example, its location requirement states that the agent must be in a location tagged `classroom` (remote learning is frowned upon in our virtual university), and the people requirement states that Norma must have a `student-of` relationship with an attendee, i.e. her professor must be present. This action increases the `accomplishment` motive by 1 and will occupy Norma for 75 minutes (or iterations) of simulation time. There are two types of actions in the Anthology system: *primary actions*, and *scheduled actions*.

Primary Actions Primary actions are atomic actions that are executed directly by a single agent. Specifically, their effects only change the motives of the agent that performs them. In Anthology, action effects are a tuple of a motive type and an integer value. When a primary action is executed, the specified motive of the agent who executed it will be adjusted by the provided integer value. Figure **??** shows an example of a primary action, `attend_class` that increases the `accomplishment` motive for any student executing it.

Scheduled Actions Scheduled actions, by contrast, are social actions involving plans undertaken by multiple agents (so they need to be "scheduled" for a time when both agents are unoccupied), and they consist of multiple primary actions. Scheduled actions affect not only the agent *instigating* the action but also *targets* other agents around them. In its current version, Anthology deems any agent in the same location as a target of a scheduled action. In lieu of effects, schedule actions have instigator and target actions which are added to the action queue for the participating agents. For instance, in a scheduled `hug_romantically` action an instigator may attempt to hug a target agent. In this case, the scheduler adds an instigating action `attempt_hug` to the instigator, and `accept_hug` or `reject_hug` onto the action queue of the target agents.

Finally, an action may be `hidden`, to represent actions that can only ever be added to the queue of an agent in response to a scheduled action. For instance, an agent may never select the action `reject_hug` without another agent attempting to hug them.

6.6.4 Execution Engine

The execution engine has two primary responsibilities: periodically decaying agent motives and triggering each agent's turn. Every time step, the execution engine calls for each agent to either make progress on an action or choose a new action to perform every turn.

When an agent is called upon to take their turn, if they have no actions remaining in their queue, they must select the next action to take. Agents are incentivized to select actions that improve their motives, so long as the requirements of an action can be fulfilled. The overall gain in motives is calculated and weighted to take into account the time it would take to complete the action and optionally any travel time required. This computed value is the utility of the action. Action utilities are then compared and the highest utility action is chosen by the agent, deferring to random chance when deciding between actions with equivalent utilities as can be seen in Listing 1.

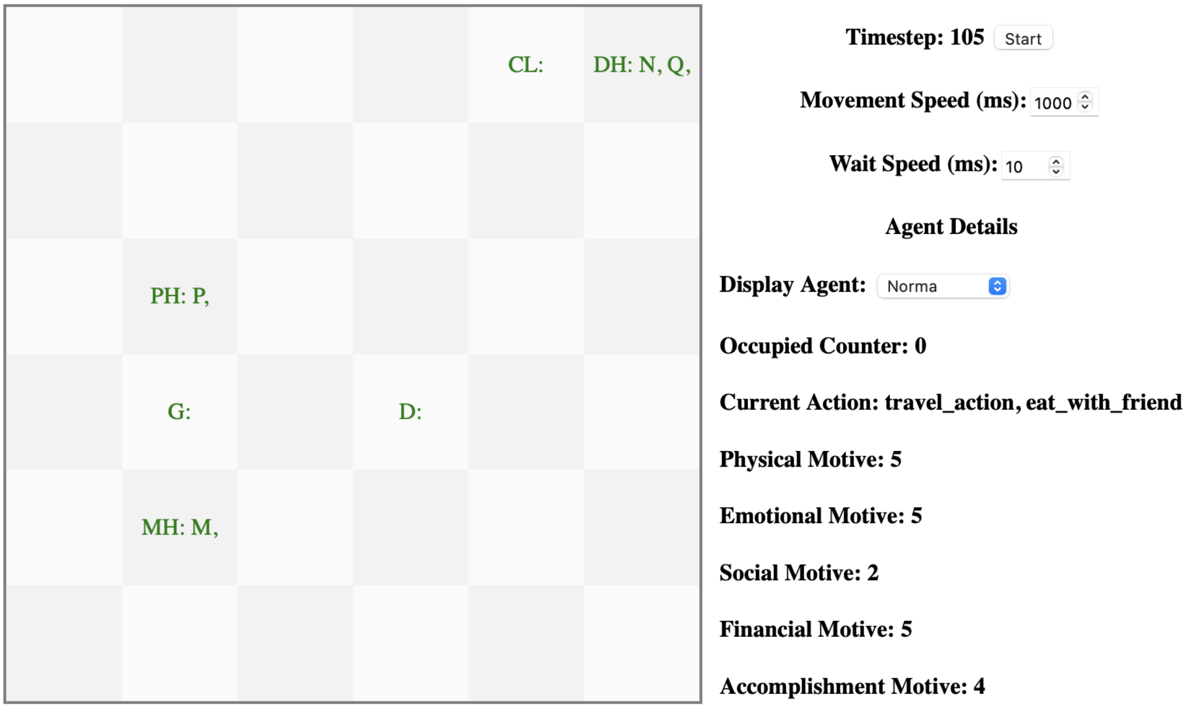


Figure 6.3: User Interface depicting our University example

6.7 User Interface

Currently, the user interface for the simulation is a configurable simple grid-based map (see Figure 6.3). Each cell in the grid can contain any number of locations and agents. Agents are denoted by capital letters representing the first letter of their name, and locations are denoted by lower case letters of the first letter of their names. There are two different running speeds for the simulation: one for when all agents are staying in the same location, presumably working on completing an action, and another for when any agent is moving about the map. This is so that users can get a better sense of how often their agents move and where to, while essentially "fast-forwarding" through idle time where all agents are stationary. Both of these speeds are measured in milliseconds and represent a delay between each time step of the execution engine. Both these speeds can be modified by users in real-time on the side panel of the user interface.

Further, users can select any agent to view real-time information about them. The side panel will display their current motive values, their action queue, and their occupied counter. This can be used to keep track of how a particular agent addresses their motives throughout the course of the simulation. To get a more fine-grain view of all agents, we

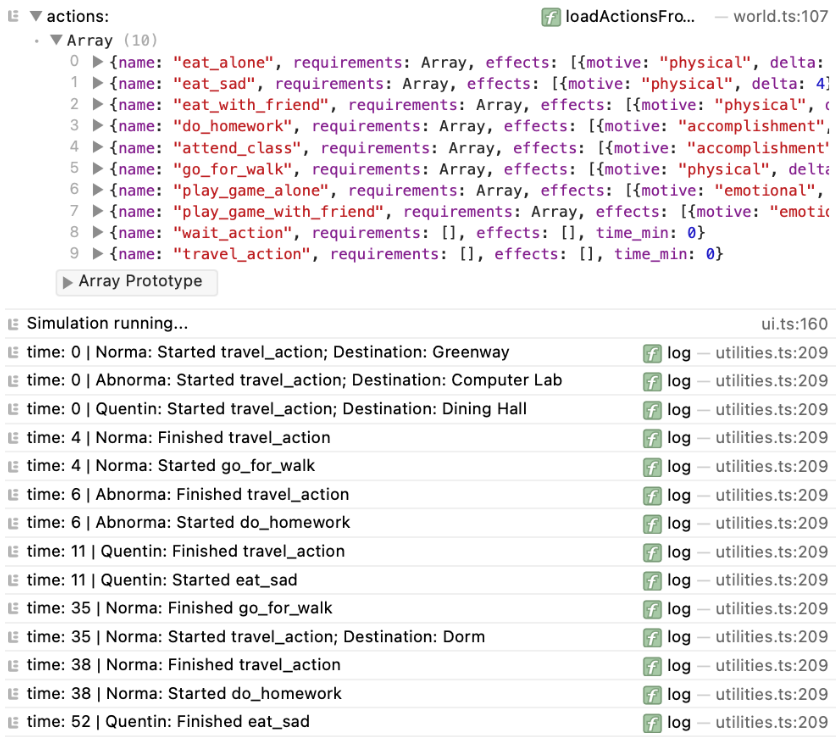


Figure 6.4: Partial Action Trace from the developer console in the browser.

also log a full action trace (see Fig. 6.4) for the simulation in the developer console of the browser.

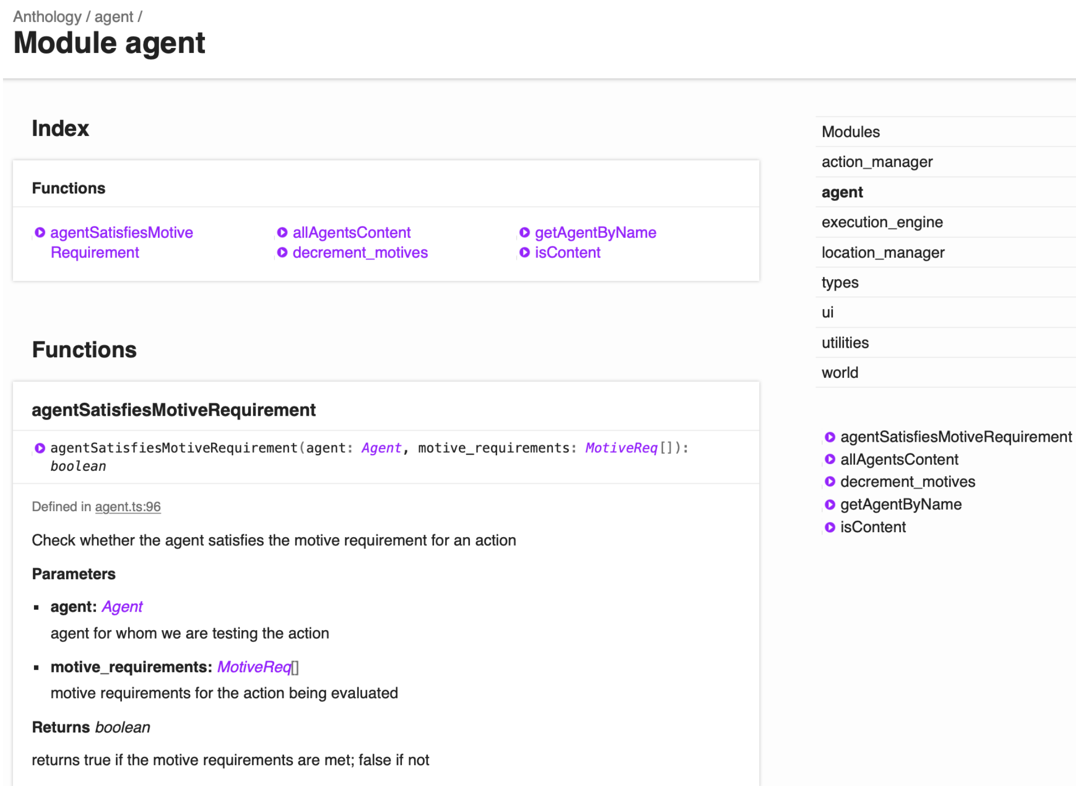
6.8 API Documentation

The Anthology system comes with a complete set of user documentation (see Figure 6.5). This documentation is auto-generated from the typescript documentation of our code base and is separated into modules for convenience.

Our documentation is detailed. We provide both a general description of every type and function, as well as what it's used for, and a description of each input and output the same.

6.9 Scalability

To measure the computational scalability of the system, data was collected on the average time of a single turn (one tick of the simulation where all agents advance their current actions or choose a new action), the total time of a complete simulation, and the total



Anthology / agent /

Module agent

Index

Functions

- agentSatisfiesMotiveRequirement
- allAgentsContent
- decrement_motives
- getAgentByName
- isContent

Functions

agentSatisfiesMotiveRequirement

```
agentSatisfiesMotiveRequirement(agent: Agent, motive_requirements: MotiveReq[]):
boolean
```

Defined in agent.ts:96

Check whether the agent satisfies the motive requirement for an action

Parameters

- **agent:** *Agent*
  agent for whom we are testing the action
- **motive_requirements:** *MotiveReq*[]
  motive requirements for the action being evaluated

Returns *boolean*

returns true if the motive requirements are met; false if not

Modules
action_manager
agent
execution_engine
location_manager
types
ui
utilities
world

- agentSatisfiesMotiveRequirement
- allAgentsContent
- decrement_motives
- getAgentByName
- isContent

Figure 6.5: View of Anthology's Documentation Generated for the Agent Module

number of turns in a complete simulation. Each of these metrics was measured as the total number of agents in the simulation was varied, and each configuration of the simulation (i.e. each trial number of agents) was tested five times and the results aggregated into a single data point for each metric. The simulation was tested with configurations of the agents ranging from 2 to 10000.

A graph detailing the average total time scaling for the simulation can be seen in in Fig. 6.6, with the average turn time following a largely similar pattern. For both metrics, notable slowdowns did not occur until 5000 agents were run simultaneously in the simulation. While turn times were all under 1ms previously, and total times ranged from 22000ms to 32000ms, at the 5000 agent mark, turn times increased to several ms and total times more than doubled. However, the number of turns remained constant regardless of the number of agents simulated.

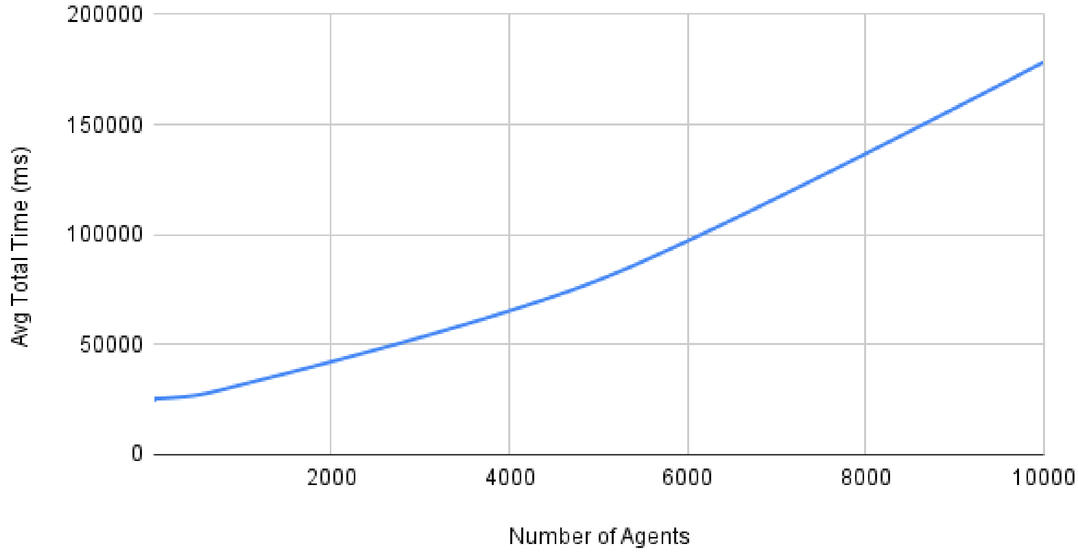


Figure 6.6: Chart depicting the average run-time when scaling from 2 to 10000 Agents

6.10 Discussion and Future Work

We posit that Anthology's design reflects the emergent nature of human social behavior. In particular, Anthology captures how simple facts about people's physical location, relationships, and motives can determine what they can, cannot, and would not do. It is our intention with this design to facilitate discussions and explorations of how people's actions are guided not just by their intentions, but also by their context in society.

We have presented Anthology's system design and its role in our ongoing research agenda to integrate social AI techniques from digital games into a usable, expressive framework for modeling and understanding human behavior. We defined our design goals from a human-centered software development perspective, explained Anthology's technical underpinnings in an implementation-independent way, and demonstrated a modeling interaction in which a scenario is constructed, run, and iterated upon to reveal simulation insights and prompt social science research questions. In the long term, we expect this project to enable reuse and reproducibility for social simulation research projects within and outside of our group, and to allow social simulation researchers to model and reason about the complex dynamics of human social behavior.

With our running example, in Fig. 6.7, we demonstrate the varied behaviors that can emerge out of Anthology by only changing the input JSON file. The ease with which the simulation can be refined, as tested during our formative evaluations in our participatory design phase, shows that Anthology's design is conducive to rapid prototyping and iteration. However, JSON files have syntax requirements that can prove to be burdensome in larger projects. These limitations may be rectified by the inclusion of a domain-specific language for Anthology akin to Kismet (Summerville and Samuel 2020).

```
time: 0 | Norma: Started travel_action; Destination: Greenway
time: 0 | Quentin: Started travel_action; Destination: Dorm
time: 4 | Norma: Finished travel_action
time: 4 | Norma: Started go_for_walk
time: 6 | Quentin: Finished travel_action
time: 6 | Quentin: Started do_homework
time: 35 | Norma: Finished go_for_walk
time: 67 | Quentin: Finished do_homework
time: 67 | Quentin: Started travel_action; Destination: Dining Hall
time: 73 | Quentin: Finished travel_action
time: 73 | Quentin: Started eat_alone
time: 99 | Norma: Started travel_action; Destination: Dining Hall
time: 105 | Norma: Finished travel_action
time: 105 | Norma: Started eat_with_friend
time: 134 | Quentin: Finished eat_alone
time: 134 | Quentin: Started eat_with_friend
```

Figure 6.7: Simulation output from our university example

Using the Participatory Research Design methods described above allowed us to iteratively design and develop the emerging design. Expert user feedback phases helped us to envision and shape the system alongside participants who will eventually form a part of our user base for the tool (Spinuzzi 2005; Mirel 1998). We have found that this perspective, allows us to leverage the intuitive and tacit knowledge of the experts in the domain, understand their priorities and needs, and develop a tool that is an extension of this knowledge. We believe that this methodology has contributed to the project making sustainable progress and that in the long term it will improve the practicability and long-term reach of the project. We will continue to iterate on the Anthology system following the same methodology and protocols. Our roadmap for future prototypes of Anthology includes (but is not limited to) some of the features listed below.

6.10.1 Utility Function:

Currently when an agent evaluates all possible actions, they will always choose the highest scoring action. This however implies that if there are two actions highly scored by our utility function, the agent will always select the higher one, and the second one will never be run. One way to solve this, would be if the agent randomly selected between the top-n actions. However, this method fails when authorial decisions imply only a few m actions are very highly scored, and the other n-m actions have low or negative utility. One way of solving this problem is by selecting the actions based on a normalized or gaussian distribution of their utility score. In that case, the highest ranked actions have a higher probability of being selected, but other actions will also have a probability based on their score.

6.10.2 Agent-specific motives:

Currently all agents have the same motives that decay at a set periodicity and increase by a set value based on the actions they are undertaking. A future version of Anthology would support agents with distinct or weighted motives. Additionally, we would like to change the factor by which the agent's motives are modified. This would allow us to model, for instance, an introverted agent that is more socially fulfilled (i.e. receives greater motive increments) by performing easier social actions (i.e. attending lab meetings), and who is easily overwhelmed by larger social actions (e.g. a crowded frat party).

6.10.3 Scheduled joint actions and interrupts:

Currently, our implementation of joint actions only supports immediate interactions between agents. A more robust implementation would allow agents to schedule joint actions for a future time. Additionally, in the current version, agents receive the entire effects (i.e. motive changes) of their actions at the end of the action duration period. This combination of features has led to hilarious (but tragic) narrative scenarios where an agent about to complete their 8-hr workday was interrupted by another agent that yelled at them, and consequently wasn't paid (i.e. received no financial or accomplishment motive gain) for the entire day.

6.11 Simulation Pains and Gains

With Anthology, our goal was primarily to create a expressive, extensible and reusable framework for social simulation. We received feedback from other researchers that appreciated how multiple formative evaluation phases had already been rolled into the existing system design as the design moved forward, improving the reusability of our framework.

During conversations with other researchers at AAAI's AIIDE conference we were applauded for our chosen methodology as it allowed us to solve a key problem they identified while designing their own simulation toolkits – that of feature creep. Researchers found it difficult to choose how to prioritize the list of features they should focus on, both within the toolkit to improve authorability, as well as those designed for the virtual characters they simulated to improve expressiveness. They found that often their choices resulted in excessive features added in to their toolkits making it too complicated or difficult to use. They appreciated our human-centered design methodology as it allowed us to prioritize features based on user requirements.

With Anthology, our formative evaluation phases were able to tackle both the main problems we had with designing Lyra. The first, that of choosing which interactions we should design, as well as the granularity of the interactions. Secondly, this time around, we were able to have a deeper evaluation than the limited one we were manuevered into with Lyra. The evaluation was able to cover not just the expressiveness of the interactions allowed by the system, but also the system itself, the utility functions we chose, and the documentation of the system.

CHAPTER

7

PRELIMINARY RESULTS

It is my position that it is time for our field to have agreed upon taxonomies and conventions that allow us to discuss both past and new social character models and frameworks. The Little Computer People Taxonomy and associated Framework could better facilitate our understanding and use of agent-based modeling in the entertainment domain and allow for better communication and collaboration among researchers and practitioners developing social simulation systems.

7.1 Informal and Early Feedback

Some informal evaluation has already been obtained on our preliminary work during published conference proceedings of the same. Within the AI Entertainment, and Social Simulation field in their work titled, *Exploring the Design Space of Social Physics Engines in Games*, Johnson-Bey et al. (2022a) describe our taxonomy as follows:

> *"[The Little Computer People Taxonomy] provides a taxonomy of social simulation in games and a shared vocabulary for comparative analysis between different social simulation systems. They are an ideal reference for thinking about design*

> *options when designing a social simulation system. Their taxonomy surfaced the following themes: communication, the flow of knowledge, relationships, and emotions."*

Within the Computational Social Science field, with their textbook for *Computational Modeling of Infectious Disease,* Von Csefalvay (2023) describe how our work was able to capture in detail a "hugely complex agent-behavioral model" and describe how it can be used to learn the *"fundamentals of agent-based, behavioral models,"* and used to *model a heterogenous population to map models of epidemics.*

7.2 Using the Little Computer People Taxonomy

This section discusses how our taxonomy can be operationalized and used to fulfill our goals to reuse, reproduce, and compare existing models. We also discuss and review our analysis, keeping in mind our initial goals. We propose that researchers referring to our taxonomy could use it in several ways.

7.2.1 As an Analytical Tool

Our taxonomy can be used to analyze existing work and set standards for evaluation based on the area of contribution. For instance, can we standardize experiment designs or target audiences that can be used to compare or evaluate physical interactions amongst social characters? What are the existing physical interactions modeled between characters that increase familial relationship valences? How does a proposed new physical interaction behavior compare? Will adding these interactions make agents more believable, and sociable, or improve communication between agents and the player?

Relatedly, the taxonomy can also be used as an object of critique in and of itself: because it encapsulates a range of choices made in a representative set of systems, we can study and identify the underlying assumptions, theories, and metaphors embedded in existing social simulation systems, and we can identify alternatives to them that are missing from this canon. Section 7.3 demonstrates this usage as a key step of the Critical Technical Practice methodology.

7.2.2 As a Design Tool

To contextualize new work, designers can enumerate their specific design choices for their system using the taxonomy. They can clarify where their contribution lies, and help situate their work with respect to previous systems. For instance, consider a narrative designer creating a simulation of the spread of political news and misinformation. Then questions they might ask of our taxonomy could be, what existing projects have designed models of knowledge, communication and relationship that allow for the deliberate spread of misinformation? What are projects that represent character knowledge and veracity of information? Are there existing relationship models that take into account trust or deceptiveness of the participant as a dimension? What communication methods would need to be used to depict truthfulness in information being dissembled.

7.2.3 As a Social Physics Engine or API

We posit our taxonomy can be used as a list of features or requirements to implement during the construction of a social physics engine, library or API. Designers could search this social physics API, for instance, for PDDL definitions of behaviours that increase the valence of romance relationships by a small amount in a newly budding relationship amongst co-workers. Thus, interactions one may not see recommended could be, "Propose having a baby." Of course, simulating such behaviours could lead to entertaining narrative tropes such as a *One Night Stand Pregnancy.*

7.2.4 For Software Engineering Processes

We propose using this taxonomy as a project planning guide. This would allow development groups to assemble team members with expertise in simulating various phenomena, such as engineers to animate characters, a developer with expertise in knowledge engineering, etc. Additionally, the taxonomy can inform architecture decisions, allowing developers to reuse components, for say, emotions, from their or other researcher's existing projects.

Additionally, our taxonomy can be used during the requirement-gathering stage in a software development project. It can dictate what features must be implemented, and prioritize them in increasing order of complexity. For instance, an initial proof of concept for an emotions model could be developed including just a binary trigger (or behavior tree) that lets characters express (e.g. using facial expressions) a type of emotion (e.g. happy) when an event occurs (for instance, at a birthday party). In a more complex model, the

Principle of Contingency could be developed, with programmatic implementation to track the antecedent cause (e.g. the event of the birthday party), and the valence in happiness (e.g. +10 to happiness), and so on for the other principles in Section 5.4.

7.3 Addressing the Unseen: Un-spun Tales

Our surveyed social simulation projects had disparate goals, ranging from pure entertainment to enabling systems understanding of real-world social phenomena. These disparate goals make the set of phenomena they choose to simulate interesting in their overlap, in our comparison of the models and the different choices their authors made for computational representation. However, equally interesting is the set of social phenomena and theories that they *didn't* simulate, and that we did not come across.

Phil Agre's theory of Critical Technical Practice (Agre 1997), and Phoebe Sengers' extension of this concept to HCI (Sengers et al. 2005), guide us to consider the underlying assumptions, values, and metaphors that simulation authors rely on, and to identify alternatives. In this set of social simulation projects, we identify the underlying assumptions and encoded social norms pertaining to (1) notions of identity; (2) the structure of human families and relationships; (3) relative power between agents; and (4) human needs and access to resources that meet those needs.

7.3.1 Identity and Social Norms

Every system we studied had a formal representation of a character or an abstraction of human identity. The data structure representing a person may contain a name, age, set of parents, personality traits, and gender, for example. The choices of data structure fields, the types of their values, and their use in the simulation all reflect assumptions about the role of identity in society. When gender is represented as a binary (Azad and Martens 2019; Ryan et al. 2015; McCoy et al. 2011a), it also enables explicit representation of sexual orientation and attraction. For instance, the game Rimworld (Studios 2018), which was not included in our study, was heavily critiqued (Lo 2016) for encoding gender-essentialist differences in the attraction models between men and women in the simulation. Gender also sometimes affects the clothing choices available for simulated characters (Miller and Summers 2007).

Apart from *The Sims* (Maxis 2003), that released an update for their most recent release that received acclaim for breaking down gender barriers (Burns 2016), the simulations we surveyed typically enforced these norms without actually *modelling* their role as norms:

they may select gendered names or pronouns (Ryan et al. 2015) for characters, or they may allow characters to have gender-neutral names, but the concept of *norm breaking* or violations of social norms is not encoded.

7.3.2 Family Structure

Simulations of family structure tend to replicate the nuclear, heteronormative, natal family: a child is born to a mother and a father, and those relationships are permanently maintained throughout the simulation (Azad and Martens 2019; McCoy et al. 2011a). Romantic relationships are typically ruled out between natal family members, leading to sometimes surprising situations for players, such as being unable to perform the Flirt action between characters who happen to be distant relatives (as seen in Section 5.3.2). These design choices are of course a simplification of real life, choosing to simulate what is considered socially normal for the target audience of players and choosing to omit social phenomena that break social norms.

As social norms in target player audiences evolve, so do the simulated phenomena of commercial social simulation games. The inclusion of same-sex relationships in *The Sims 4* and *Animal Crossing*, for example, has been a site of controversy (Kelleher et al. 2020).

7.3.3 Power Relationships

Of the simulations we surveyed, few represent power relationships between characters (Meehan 1977; Si et al. 2005). In general, edges in the relationship graph are assumed to be *consensual*, even if not reciprocal, and each character has the same set of actions and constraints as every other character. Power imbalances such as the privilege afforded racial and gender bias are not simulated, nor are coercive relationships that in real life may emerge between someone dependent on another for necessary resources (Meehan 1977), such as an employee and their boss (Ryan et al. 2015) or a child and a bully (Si et al. 2005). Generally, the only sources of conflict between characters are due to one-on-one interactions, such as bullying and insulting, rather than systemic, trans-generational power imbalances between groups.

7.3.4 Meeting Needs

Power imbalance, described above, is strongly related to the idea that some people may depend on others to meet their basic needs. In American capitalism, people need to earn money through work to obtain food, shelter, and clothing. In reality, jobs may not be available, a person may not possess the skills needed to get a job, or the money earned from a job may not be sufficient to meet the living standards. In the simulations we surveyed, basic survival needs (if simulated) are generally *possible* to meet through non-coercive means. It is always possible for a character to find work, and in some cases, they are given jobs as favours by family members if they need one (Ryan et al. 2015). Of course, the real-life dynamics of labour and pay are much more complicated. Some playable simulations exist as effective critiques of capitalism's exploitation of labour, such as *Papers, Please* (Morrissette 2017), and *To Build a Better Mousetrap* (Molleindustria 2014). However, these games simulate economic phenomena in isolation from their effects on social interaction.

7.3.5 Modeling Social Norms

A frequently observed pattern in all of these examples is that simulations may *encode* certain social norms (e.g. programmatically enforcing or creating different responses to socially-abnormal actions), but they do not model the formation, absorption, perpetuation, and transformation of those norms. Some academic projects are attempting to develop richer models of social norms and propriety in other simulations (Blass and Horswill 2015); however, these models have not yet been realized in playable experiences.

7.4 Does the Little Computer People Taxonomy and Framework support our goals?

The goals of developing a taxonomy and a vocabulary were to enable researchers and industry designers building social agent models to communicate effectively, reuse one another's work, and compare projects across a wide variety of narrative domains. Here we review the outcome of our analysis in light of our goals.

- We conducted a preliminary investigation into the need for NPC model-based analysis
- We conducted an in-depth investigation of a few selected models from research and industry and were able to:
  - Identify the commonalities and differences in the various systems in terms of the level of granularity, social state assumptions, communication, the flow of knowledge, social relationships and behaviours

- Identify differences in implementation: where systems shared similar “verbs” but implemented them according to different models or assumptions
- Discuss how this taxonomy could be used to compare play experiences or research methods or goals

CHAPTER

8

PROPOSED WORK

As discussed in Chapter 1.4 my work has five research thrusts. To describe my proposed work in this document, I first briefly outline the three research thrusts that have already been completed, the work in progress, and finally, what remains to be done. I will describe each task in further detail and then summarize my work plan with a Gannt Chart that describes my path from here to my defense.

8.1 Research Status

My initial attempts at Social Simulation with the US-Voting simulation project (described Chapter 3) that eventually turned into Lyra introduced me to some of the common problems social simulation researchers face. The original simulation has been described in detail and published at the 2018 Experimental AI in Games Workshop at the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (Azad and Martens 2018). I followed up this work with a human subject study to analyze Lyra's effectiveness at believably simulating opinion propagation and the social dynamics of politically charged conversations. This work was published in the proceedings of The 2019 AAAI conference on Artificial Intelligence and Interactive Digital Entertainment (Azad and Martens 2019).

These initial social simulation attempts helped me uncover problems faced by experienced social simulation and game researchers when designing social characters for virtual worlds through an in-depth survey of both commercial and research-based social simulation works in the field as described in Chapter 4. I discovered researchers used disparate approaches to describe their simulations and research advances. Researchers either (1) used similar terms to describe wildly different phenomena, or (2) used conflicting terms to describe the same phenomenon, or (3) could agree on terminology. However, they modeled the underlying social phenomenon with varying granularity and rulesets while using the same terminology. This lack of consensus made it challenging to identify, evaluate, classify, or build on existing knowledge. To address this lack of consensus and improve the reuse, reproducibility, and comparability of social simulation research, my prior work designed the Little Computer People Taxonomy as described in Chapter 5. The Little Computer People Taxonomy aims to resolve the observed inconsistencies in the literature and methodologies for researching and developing social simulations. This work resulted in a journal paper presented at CHI Play 2021 in the Proceedings of the ACM on Human-Computer Interaction (PACMHCI) Journal (Azad and Martens 2021).

Learning from the research with Lyra (described in Chapter 3), I designed Anthology (described in Chapter 6) with a goal to create a social simulation authoring tool that was both usable and expressive. Our human-centric design methodology for designing Anthology, along with its system description, user interface, API documentation and an early analysis of Anthology's scalability has been documented. The resulting work was published at The 2022 AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (Azad et al. 2022).

Designing these two simulations allowed me to both gain clarity on methods by which social simulations are created. I was able to understand how the taxonomy and it's accompanying framework could support the difficulties that designing and evaluating such a system could faced.

The proposed work of this dissertation is in two parts. In the next sections, I discuss my proposed work towards Research Thrust #4 in developing the social character framework and authoring library, and Research Thrust #5 towards an evaluation of the impact of the Little Computer People taxonomy and proposed framework on experienced researchers and users in the social simulation games and research field.

8.2 Developing the Framework

For Research Thrust #3 I proposed the design and development of my second research artefact, a "Social Physics Framework" (McCoy et al. 2011a). The development of this tool would address RQ3 (repeated for convenience):

> ***RQ 3:*** *How can we operationalize the designed taxonomy into a framework that the narrative and entertainment intelligence community can use?*

I have identified and created all the components required to operationalize the taxonomy into the Social Physics Framework across my earlier works. I propose that a modular, scalable, flexible Social Physics Framework that answers RQ3 requires the following necessary components:

An Agent-Based Social Simulation Base: We require a base simulation framework that is capable of simulating a multi-agent system, with autonomous agents capable of interacting with one another at varying levels of granularity. This component has been designed and developed in the Lyra, Anthology, and Clockwork systems.

A Discrete Event Simulation: I posit we require a Discrete Event Simulation system to work in conjunction with the Agent-Based Social Simulation base in order to allow for better scalability, with centralized scheduling and control of events. I have designed, developed, and tested this component in the Clockwork system.

An Agent and Environment Model: A model of common and specific agents, a representation of the environment, and methods of interaction between them. I have developed varying agent models for the Lyra, Anthology, and IBM-Clockwork simulations and discuss other existing social agent models surveyed in Chapter 4. I propose to use the LCP Taxonomy to design and prioritize requirements for the social agent model for this dissertation. I anticipate each theme of the taxonomy to be a modular library that can be plugged into the base simulation. Alternatively, my targeted user groups can swap out the taxonomy modules for their own models or utility functions to compare and contrast simulation results and further research.

An Authoring Toolkit: The design of a usable, flexible authoring toolkit that can be used to author agent models and allow social simulation researchers to easily simulate social agents and study their complex underlying social phenomenons. I have designed an authoring toolkit with my prior work with Anthology. I plan to reuse components from Anthology to create the final Social Physics Framework.

The Social Physics Framework is a significant software engineering effort. I anticipate that the initial feedback I obtain from expert users will help direct the design choices of the project. The proposed tasks have been described in further detail below.

8.2.1 Creating The Baseline Simulation

I plan to reuse the architecture and code from all my previous simulations, Anthology (see Chapter 6), Lyra (see Chapter 3) and the IBM COVID-Simulation created by me. I will refactor the simulations with the lessons learnt by me to design the baseline simulation.

Problem: Scalability

One problem I will tackle with the base simulation is the problem of achieving scalability without compromising granularity and control in the simulation. This is a problem that I noticed both during my in-depth survey of social simulations. For instance, existing work by Ryan et al. (2016b) describes their approach to solving this scalability problem: by extrapolation. Rather than simulate every timestep during world generation, in Talk of the Town (Ryan et al. 2015), the authors describe their dwarf-like (Adams and Adams 2006) simulation, iterating through a handful of timesteps a game year and extrapolate charge and spark increases according to the amount of time since the last simulated timestep. This extrapolation provides us with more efficient computation. However, this extrapolation comes at the cost of the granularity and accuracy of the simulation.

Proposed Approach: Scalability

I propose using Discrete Event Simulation methods in conjunction with the Artificial Intelligence decision-making algorithms used commonly in social simulations to improve the scalability of the simulation without compromising on granularity. The result of this will be a scalable, baseline authoring framework.

With Anthology, I have designed a human-centric, extensible, expressive and reusable ABSS toolkit in Typescript. With more recent work, at IBM Research I successfully researched, designed, experimented with and developed a simulation integrating the DES and ABSS methodologies together into a simulation using the DES SimPy framework [1]. The simulation was lauded during the research poster demonstration for the same for the ability of the

[1] The associated paper is still under construction. The project was done as a summer project at IBM Research.

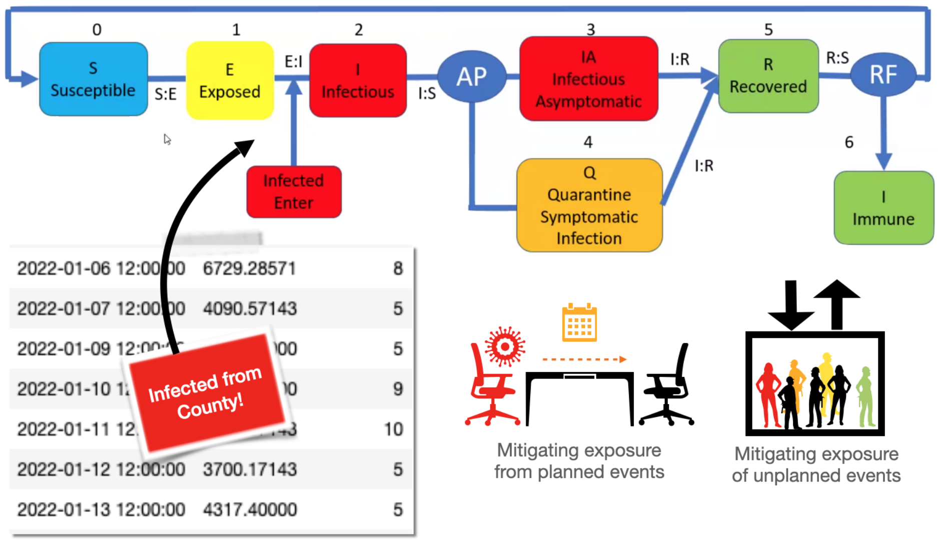


Figure 8.1: An advance look at the ABSS+DES COVID-10 Simulation designed and developed for IBM Research. ABSS Digital Twins of employees allowed us to design and account for employee preferences (eg. on masking, or social distancing), or take into account employee relationship dynamics. The DES system allowed us to schedule and simulate COVID SEIR events, perform Planning-On-Demand, and handling the resource bottlenecks for the system as a whole.

simulation to take into account individual agent desires, motivations, and preferences (using primarily ABSS techniques) and integrate with that the CDC SEIR model and make it into a scalable simulation with detailed granularity (using DES methodologies).

My proposed work will integrate these two methodologies and approaches into a single framework, The Little Computer People Framework.

Anticipated Challenges: Scalability

I am currently experimenting with re-architecturing our Typescript Skeletal framework from Anthology and adding in my ABSS+DES SimPy codebase. I plan to build the base simulation using Typescript's DES library, SimScript. However, the re-architecturing of the framework is a significant software development effort, and may have unforeseen delays. This could extend the base simulation code timeline I outline in Table 8.2.

Problem: Designing a Usable, Flexible framework

As we see in 2, there have been attempts at creating a modular framework before. These attempts have largely been focused on making an existing research framework (for instance, that of FaTiMa (Dias et al. 2014) or CiF (McCoy et al. 2010a)) more accessible, and thus easier to use by the community. These efforts differ from ours since we aim to make an even more generic baseline simulation that can be used for multiple differing environments (whether a prom week, a town, a workplace, or otherwise), and at varying levels of granularity of behavior.

Proposed Approach: Designing a Usable, Flexible framework

I aim to use the human subject design-focused methodology that was so successful in Anthology (see Section 6.4). I expect to continue to design and develop the system using Participatory Research Design methodology to run the human subject studies using our existing IRB for the social simulation authoring tool. With our approach, participants are asked to construct any scenario or environment they can imagine so we can learn the capabilities and restrictions of using our framework.

I believe this methodology will help to resolve the problem of feature creep, and to ensure we end up with a usable, and flexible framework. Further formative evaluations have been discussed in the section 8.3 below. We plan to use periodic and continuous evaluations to inform our framework architecture decisions.

Anticipated Challenges: Designing a Usable, Flexible framework

Running user studies for the participatory design approach is time-consuming. Even if we aim to study 1-2 users per iteration, the design process can grow extensive, and take up more of the timeline than currently accounted for.

On the other hand, by designing and developing the framework without the Participatory Design methodology, we risk creating a framework that has the same issue as other simulations. That is, they do not meet our re-usability or research collaboration research goals. Thus, the design, development, and evaluation methodologies we use need to be carefully balanced.

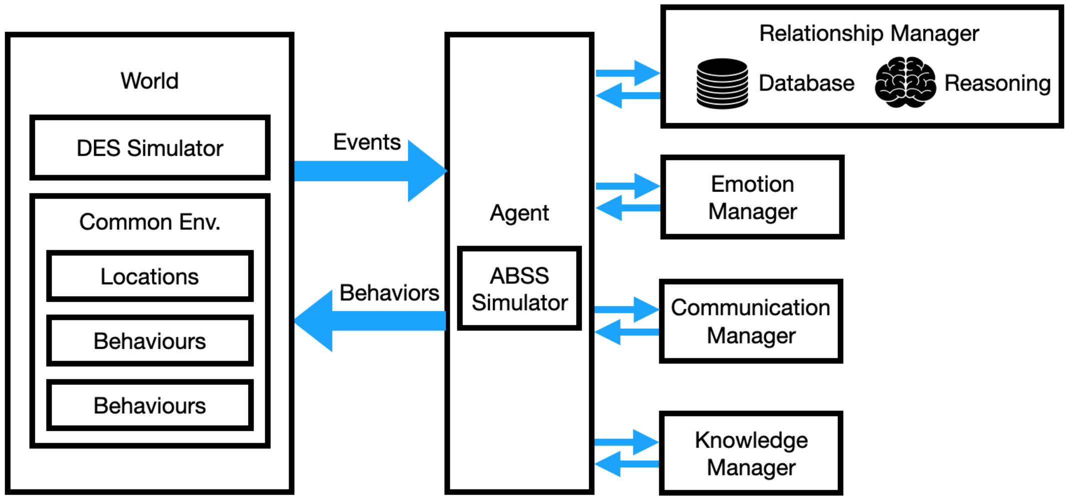


Figure 8.2: Proposed Architecture for The Little Computer People Framework

8.2.2 Developing the Taxonomy Framework

Once the baseline simulation framework is completed, I will be designing library modules for the LCP Taxonomy themes (for instance, for relationships) in Typescript.

Proposed Approach

My goal is for the designed modules to closely match the Little Computer People Taxonomy (see Chapter 5) with appropriate scaffolding of the code to create models and controllers (modeling an MVC architecture) that the designers can use. These modules can be optionally added into the baseline simulation using simple import statements. The modules will extend the capabilities of the simulation and the agents simulated. An example of what this architecture could look like has been shown in Fig. 8.2.

Relationships: The most straightforward theme to design and add into the simulation would be the Relationship theme. Relationships have been implemented by (almost) every social simulation to some extent or another. The relationship types, attributes, dimensions and dynamics are clearly defined, and their rules and constraints can be implemented in the simulation.

Communication and Flow of Knowledge: For the Communication and Flow of Knowledge themes, we currently plan to demo how the Lyra discussion module could be used in combination with the base simulation. A Web API for Lyra is being developed. We plan to allow for an application to register their characters with the API. Next, applications can trigger the Lyra discussion mode, or contemplation mode for any subset of characters. The API will contain a database that tracks historical character opinion/knowledge changes, and returns the latest character views to the simulation.

Emotions: Since my prior work has not included Emotional modeling research, I do not currently plan to model emotions in the simulation engine.

Anticipated Challenges:

I have previously created a Relationship component (in Anthology, and Lyra), a Communication component (in Anthology and Clockwork), a Flow of Knowledge component (in Lyra). While I do not anticipate any problems during the implementation of the themes above, as mentioned in Chapter 4, most surveyed artifacts have not yet been modular enough to be easily reproduced or reused within the context of another simulation (agent, or environment). Thus, while these components have been created by me individually, this would be the first time I unite them in a single simulation. Errors caused due to this may affect my development timeline.

8.3 Evaluating the Impact

I propose two evaluations for my dissertation, to evaluate both the Little Computer People taxonomy, and the resulting simulation framework. The taxonomy has already been designed and is completed. As such, the evaluation of the taxonomy will take place in parallel with the development of the framework. The framework itself is being developed using Participatory Design Research Methodology, and so multiple formative evaluation phases will be conducted during the system design and prototyping of the framework. We will also aim to conduct a summative evaluation.

Thus, for my fifth thrust *(in progress)*, I propose to evaluate the impact of the taxonomy and framework on experienced researchers and game designers in terms of its applicability to their work and goals. Towards this end, I have outlined the research question associated with my evaluation.

RQ 4: *What is the impact of the taxonomy and framework on both experienced social simulation researchers and game designers in terms of its applicability to their modeling* *process? Can the taxonomy and framework be used to evaluate existing social simulations by users of the simulations?*

I expect the formal evaluation and human subject study to give us further insight into how my work further impacts researchers and users of the taxonomy and framework, as well it's applicability to their own work. Finally, prior work has established how it is difficult to evaluate and compare different social simulations (Johnson-Bey et al. 2022a; Azad and Martens 2021). I plan to develop a evaluation that uses the LCP taxonomy to model a survey that asks developers and designers to rate their own social simulations against the themes and dimensions of the taxonomy. We will also ask participants to rate other social simulations on the same scale. Comparing simulations are rated on the taxonomy scale by their own desigenrs and users would allow us a further method to compare and evaluate simulations. We can learn how the simulations differ both in the overlap of these ratings (see Section 7.2: Using the Taxonomy), but also gain insights from the themes that aren't simulated or missing (see Section 7.3: Un-spun Tales). The evaluation will also be designed to measure the usability and expressivity of the taxonomy and framework using vocabulary from the Cognitive Dimensions framework (Blackwell and Green 2000).

Survey Questionnaire

At this point, the surveys have not been fully designed. However, my current expectation is to ask the following questions including (but not limited to):

- Pre-Survey Data:
  - Demographics
  - Experience with Social Simulation Design
  - Experience with Social Simulation Development
- During the Survey: For each theme overviewed, we will ask participants questions to clarify the sub-questions listed above. For instance, to answer SQ 4.1, we will ask
  - Rate the clarity of the theme, or to what extent did the participant find the understandable? (Likert + Conversational data).

- To what extent did the theme make the underlying social theories modelled more clear or accessible? (Likert + Conversational Data)
- How applicable is the theme to the work of the participant? (Likert + Conversational data)

- To test the impact on communication of existing research:
  - How well is the taxonomy capable of representing or modeling an existing simulation designed by the participant? (LIKERT)
  - Can they describe their social simulations using the taxonomy? (Checkboxes + Conversational Data)
  - What is the taxonomy not able to capture when describing their own simulation? (Unstructured Text + Conversational Data)
- To test the impact on comparison of simulations we ask the participant to Describe a simulation they have not designed using the taxonomy. What is the taxonomy able to capture about the model? What is the taxonomy not able to capture about the model? (Check Boxes, Conversational/Unstructured Text data)
- Post Survey Questions:
  - What themes or sub-themes did the taxonomy help the participant discover?
  - How useful would having the taxonomy be to the process of (1) Requirement Gathering, (2) as a Design Tool, (3) As a Social Physics Engine, (4) as an Analytical Tool? (Likert + Conversational data)

Measures:

I will collect two source of data, audio recordings of conversations with the participant, and survey data, both structured and unstructured. The audio recordings will be used only to transcribe conversations into text that can be qualitatively analyzed later and then discarded. These data sources are intended to test the following focused sub-questions:

SQ 4.1: What is the impact on expert researchers from the domain in terms of (1) Clarity of the taxonomy, and (2) Effectiveness of the classification in making the underlying social theories more accessible?

SQ 4.2: What is the impact of the taxonomy and framework on experienced researchers in terms of its applicability to their work? What barriers, if any, were removed by users utilizing the Little Computer People Taxonomy or Framework in terms of its applicability to their work? What new barriers, if any, were introduced?

SQ 4.3: What is the impact of the taxonomy and framework on experienced researchers in the community that undertake social simulations with their goals to (1) Build and design new social character models, (2) Reproduce or evaluate results from existing social simulation work, (3) Compare existing social simulation systems, and (4) Collaborate with other researchers in the domain

SQ 4.4: What is the impact of the taxonomy and framework on the democratization of character authoring tools? Does using the taxonomy or framework make it easier or hinder existing researchers or game developers in the field with their goals of designing social simulations?

8.4 Research Timeline

I aim to defend my dissertation by the middle of Spring 2024. I've included a list of tasks in the proposed timeline in Table 8.2, and a Gannt Chart in Fig. 8.3 below for more clarity.

These tasks further establish my publication goals for the year. However, I anticipate some of the later publications to have submission/acceptance periods that conclude after my defense.

Table 8.2: Proposed Timeline

Task Type	Task Name	Start Date	End Date
Coding #1	Creating the Base Simulation	Feb 1, 2023	May 1, 2023
Writing	IBM Paper: ABSS+DES	Mar 1, 2023	Apr 30, 2023
Writing	IBM Paper: Epidemic Simulations	Apr 15, 2023	Jul 1, 2023
Coding #2	Taxonomy Framework Implementation	Apr 1, 2023	Aug 1, 2023
Evaluation #1	LCP Taxonomy - Design + IRB	Mar 30, 2023	May 1, 2023
Writing	Anthology: Base Simulation Paper	Apr 1, 2023	May 26, 2023
Evaluation #1	LCP Taxonomy - Survey	May 1, 2023	Jun 15, 2023
Evaluation #1	LCP Taxonomy - Analysis	Jun 15, 2023	Jul 31, 2023
Writing	LCP Taxonomy Paper	Jun 1, 2023	Sep 15, 2023
Evaluation #2	LCP Framework - Design + IRB	Aug 1, 2023	Sep 15, 2023
Evaluation #2	LCP Framework - Survey	Sep 15, 2023	Nov 1, 2023
Evaluation #2	LCP Taxonomy - Analysis	Nov 1, 2023	Nov 30, 2023
Writing	LCP Framework Paper	Oct 1, 2023	Dec 15, 2023
Writing	Dissertation	Oct 30, 2023	Jan 30, 2024

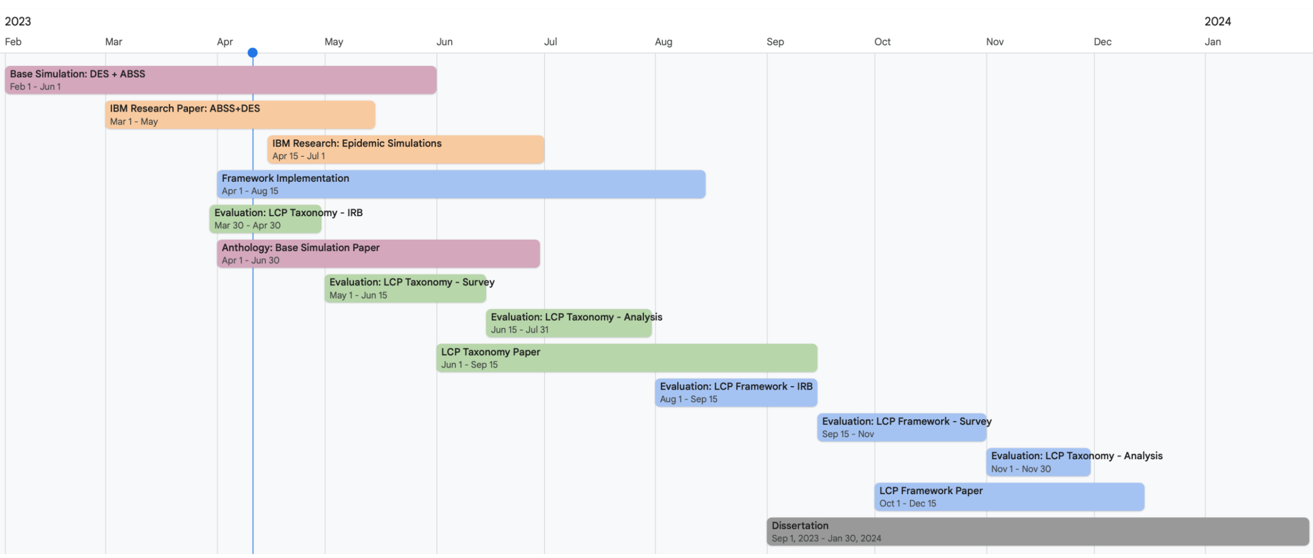


Figure 8.3: Gantt Chart for the Timeline above

REFERENCES

Adams, T. and Adams, Z. (2006). Slaves to armok: God of blood chapter ii: Dwarf fortress. *PC Game). Bay*, 12.

Afonso, N. and Prada, R. (2008). Agents that relate: Improving the social believability of non-player characters in role-playing games. In *International Conference on Entertainment Computing*, pages 34–45. Springer.

Agre, P. (1997). Computing as a social practice. *Reinventing Technology, Rediscovering Community, P. Agre and D. Schuler,(eds)*.

Allport, G. W. (1968). The historical background of modern social psychology. *Handbook of social psychology*.

AllSides (2018). Balanced news via media bias ratings for an unbiased perspective.

Asch, S. E. (1955). Opinions and social pressure. *Readings about the social animal*, 193:17–26.

Azad, S. (To Appear). To appear. In *To Appear*, volume To Appear, page To Appear.

Azad, S. and Martens, C. (2018). Addressing the elephant in the room: Opinionated virtual characters. In *AIIDE Workshops*.

Azad, S. and Martens, C. (2019). Lyra: Simulating believable opinionated virtual characters. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 15, pages 108–115.

Azad, S. and Martens, C. (2021). Little computer people: A survey and taxonomy of simulated models of social interaction. *CHI Play 2021 - The Proceedings of the ACM on Human-Computer Interaction (PACMHCI) Journal*, 5(CHI PLAY):30.

Azad, S., Wellnitz, J., Garcia, L., and Martens, C. (2022). Anthology: A social simulation framework. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 18(1).

Bateman, C. and Boon, R. (2005). *21st Century Game Design (Game Development Series)*. Charles River Media, Inc.

Bates, J. et al. (1994). The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125.

Benioff, D. and Weiss, D. B. (2019). The iron throne.

Blackwell, A. F. and Green, T. R. (2000). A cognitive dimensions questionnaire optimised for users. In *PPIG*, volume 13. Citeseer.

Blass, J. and Horswill, I. (2015). Implementing injunctive social norms using defeasible reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 11.

Brassel, K.-H., Möhring, M., Schumacher, E., and Troitzsch, K. G. (1997). Can agents cover all the world? In *Simulating social phenomena*, pages 55–72. Springer.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101.

Brichacek, A. (2016). Six ways the media influence elections. *University of Oregon School of Journalism and Communication (Online News Page).*

Budak, C., Goel, S., and Rao, J. M. (2016). Fair and balanced? quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1):250–271.

Burns, J. (2016). 'the sims' drops gender barriers, extending freedom of self-expression by avatar.

Castelfranchi, C., de Rosis, F., and Falcone, R. (1997). Social attitudes and personalities in agents. In *Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents.*

Cavazza, M., Charles, F., and Mead, S. J. (2002). Interacting with virtual characters in interactive storytelling. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1*, pages 318–325. ACM.

Cavazza, M. and Pizzi, D. (2006). Narratology for interactive storytelling: A critical introduction. In *International Conference on Technologies for Interactive Digital Storytelling and Entertainment*, pages 72–83. Springer.

Cham, J. (2009). Brain on a stick.

Champandard, A. J. (2003). *AI game development: Synthetic creatures with learning and reactive behaviors.* New Riders.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Davidsson, P. (2002). Agent based social simulation: A computer science view. *Journal of artificial societies and social simulation*, 5(1).

Dias, J., Mascarenhas, S., and Paiva, A. (2014). Fatima modular: Towards an agent architecture with a generic appraisal framework. *Emotion modeling: Towards pragmatic computational models of affective processes*, pages 44–56.

Dias, J. and Paiva, A. (2005). Feeling and reasoning: A computational model for emotional characters. In *Portuguese conference on artificial intelligence*, pages 127–140. Springer.

Dickinson, M. L., Wardrip-Fruin, N., and Mateas, M. (2017). Social simulation for social justice. In *Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Dow, S., Mehta, M., Lausier, A., MacIntyre, B., and Mateas, M. (2006). Initial lessons from ar façade, an interactive augmented reality drama. In *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*, page 28. ACM.

Durlauf, S. N. and Young, H. P. (2001). The new social economics. *Social dynamics*, pages 1–14.

Eger, M. (2020). Murder mysteries: The white whale of narrative generation? In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 16, pages 210–216.

Eger, M. and Mathewson, K. W. (2018). dairector: Automatic story beat generation through knowledge synthesis. *In Proceedings Joint Workshop on Intelligent Narrative Technologies and Intelligent Cinematography and Editing (INT-WICED) at AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE'18). Edmonton, Alberta, Canada.*

Ekman, P. (1992). Are there basic emotions? *Psychological review*, 99:550–553.

Electronic Arts, S. (2000). The sims - stategywiki.

Entman, R. M. (2007). Framing bias: Media in the distribution of power. *Journal of communication*, 57(1):163–173.

Evans, R. and Short, E. (2014). Versu—a simulationist storytelling system. *IEEE Transactions on Computational Intelligence and AI in Games*, 6(2):113–130.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.

Fowler, S. M. and Pusch, M. D. (2010). Intercultural simulation games: A review (of the united states and beyond). *Simulation & Gaming*, 41(1):94–115.

Franklin, S. and Graesser, A. (1997). Is it an agent, or just a program? procs. ecai'96 workshop on agent theories, architectures, and languages.

Gervás, P. (2009). Computational approaches to storytelling and creativity. *AI Magazine*, 30(3):49–49.

Goffman, E. et al. (2002). The presentation of self in everyday life. 1959. *Garden City, NY*, 259.

Gratch, J. and Marsella, S. (2005). Evaluating a computational model of emotion. *Autonomous Agents and Multi-Agent Systems*, 11(1):23–43.

Gray, J. (2020). The sims and animal crossing rise among amazon's best selling products as people simulate a life outside of quarantine.

Grey, C. (2011). Politics in the animal kingdom.

Grimm, V. and Railsback, S. F. (2005). *Individual-based modeling and ecology*. Princeton university press.

Guimaraes, M., Santos, P., and Jhala, A. (2017). Cif-ck: An architecture for social npcs in commercial games. In *2017 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 126–133. IEEE.

Guo, A. (2020).

Hales, D., Rouchier, J., and Edmonds, B. (2003). Model-to-model analysis. *Journal of Artificial Societies and Social Simulation*, 6(4).

Hayes-Roth, B. and Doyle, P. (1998). Animate characters. *Autonomous agents and multi-agent systems*, 1(2):195–230.

Hernandez, P. (2015). The sims players confess their most evil deeds.

Hickey, G. and Kipping, C. (1996). A multi-stage approach to the coding of data from open-ended questions. *Nurse researcher*, 4(1):81–91.

Ho, D. Y. (1998). Interpersonal relationships and relationship dominance: An analysis based on methodological relationism. *Asian Journal of Social Psychology*, 1(1):1–16.

IBM Research (2018). Ai and bias - ibm research.

Isbister, K. and Doyle, P. (2002). Design and evaluation of embodied conversational agents: A proposed taxonomy. In *The first international joint conference on autonomous agents & multi-agent systems*.

Jenks, G. F. (1967). The data model concept in statistical mapping. *International yearbook of cartography*, 7:186–190.

Johnson, W. L., Marsella, S., and Vilhjalmsson, H. (2004). The darwars tactical language training system. In *Proceedings of I/ITSEC*.

Johnson-Bey, S., Nelson, M. J., and Mateas, M. (2022a). Exploring the design space of social physics engines in games. In *International Conference on Interactive Digital Storytelling*, pages 559–576. Springer.

Johnson-Bey, S., Nelson, M. J., and Mateas, M. (2022b). Neighborly: A sandbox for simulation-based emergent narrative. In *2022 IEEE Conference on Games (CoG)*, pages 425–432. IEEE.

Jonasson, J., Lech, B., Azad, S., and Martens, C. (2021). Intertwining world and narrative generation for a mobile role-playing game.

Jowett, L. (2014). The girls who waited? female companions and gender in doctor who. *Critical Studies in Television*, 9(1):77–94.

Kelleher, P., Baska, M., and Nightingale, E. (2020). Nintendo shuts down rumours two animal crossing characters are in a gay relationship – but fans simply aren't having it.

Kelly, H. (2019). It's pretty creepy to root for game of thrones incest, right?

Kirilenko, A. P. and Stepchenkova, S. (2016). Inter-coder agreement in one-to-many classification: fuzzy kappa. *PloS one*, 11(3):e0149787.

Kreminski, M., Dickinson, M., and Wardrip-Fruin, N. (2019). Felt: a simple story sifter. In *International Conference on Interactive Digital Storytelling*, pages 267–281. Springer.

Kugele, S. and Franklin, S. (2013). General intelligence requires autonomous, cognitive, intentional agents. *Advances in Cognitive Systems*.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, 108(3):480.

Kuppens, P. and Verduyn, P. (2017). Emotion dynamics. *Current Opinion in Psychology*, 17:22–26.

Kybartas, B. and Bidarra, R. (2016). A survey on story generation techniques for authoring computational narratives. *IEEE Transactions on Computational Intelligence and AI in Games*, 9(3):239–253.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford university press.

Lech, B., Azad, S., Welnitz, J., Jonasson, J., and Martens, C. (2021). Designing a combined world and story procedural content generation engine.

Lee, H. (1960). *To Kill a Mockingbird*. J. B. Lippincott & Co.

Leeper, T. J. and Slothuus, R. (2014). Political parties, motivated reasoning, and public opinion formation. *Political Psychology*, 35:129–156.

Leino, O., Wirman, H., and Fernandez, A. (2008). *Extending experiences: structure, analysis and design of computer game player experience*. Lapland University Press.

Lewin, K. (1951). Field theory in social science: selected theoretical papers (edited by dorwin cartwright.). *American Psychological Association*.

Lim, M. Y., Dias, J., Aylett, R., and Paiva, A. (2008). Improving adaptiveness in autonomous characters. In *International Workshop on Intelligent Virtual Agents*, pages 348–355. Springer.

Lisetti, C. L. (2002). Personality, affect and emotion taxonomy for socially intelligent agents. In *FLAIRS Conference*, pages 397–401.

Lo, C. (2016). How rimworld's code defines strict gender roles.

Loper, E. and Bird, S. (2002). Nltk: the natural language toolkit. *arXiv preprint cs/0205028*.

LTD, C. (2017). Stardew valley.

Luo, L., Cai, W., Zhou, S., Lees, M., and Yin, H. (2015). A review of interactive narrative systems and technologies: a training perspective. *Simulation*, 91(2):126–147.

Lyon, S. (2020). Why "animal crossing" wins over "the sims" every time.

Macvean, A., Hajarnis, S., Headrick, B., Ferguson, A., Barve, C., Karnik, D., and Riedl, M. O. (2011). Wequest: scalable alternate reality games through end-user content authoring. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, page 22. ACM.

Magerko, B., Manzoul, W., Riedl, M., Baumer, A., Fuller, D., Luther, K., and Pearce, C. (2009). An empirical study of cognition and theatrical improvisation. In *Proceedings of the seventh ACM conference on Creativity and cognition*, pages 117–126.

Maidstone, R. (2012). Discrete event simulation, system dynamics and agent based simulation: Discussion and comparison. *System*, 1(6):1–6.

Majid, M. A. and Herawan, T. (2013). Modelling reactive and proactive behaviour in simulation: A case study in a university organisation. *International Journal of Multimedia and Ubiquitous Engineering*, 8(6):329–338.

Mao, W. and Gratch, J. (2004). Social judgment in multiagent interactions. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 210–217.

Marsella, S., Gratch, J., Petta, P., et al. (2010). Computational models of emotion. *A Blueprint for Affective Computing-A sourcebook and manual*, 11(1):21–46.

Marsella, S. C., Pynadath, D. V., and Read, S. J. (2004). Psychsim: Agent-based modeling of social interactions and influence. In *Proceedings of the international conference on cognitive modeling*, volume 36, pages 243–248.

Martens, C., Iqbal, O., Azad, S., and Ingling, M. (2018). Villanelle: Towards authorable autonomous characters in interactive narrative. In *2018 AIIDE Workshops*.

Maslow, A. H. (1943). A theory of human motivation. *Psychological review*, 50(4):370.

Mateas, M. and Stern, A. (2003). Façade: An experiment in building a fully-realized interactive drama. In *Game developers conference*, volume 2, pages 4–8.

Maxis, E. (2003). The sims.

Mayring, P. (2004). Qualitative content analysis. *A companion to qualitative research*, 1:159–176.

McCoy, J., Treanor, M., Samuel, B., Mateas, M., and Wardrip-Fruin, N. (2011a). Prom week: social physics as gameplay. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, pages 319–321. ACM.

McCoy, J., Treanor, M., Samuel, B., Reed, A. A., Wardrip-Fruin, N., and Mateas, M. (2012). Prom week. In *Proceedings of the International Conference on the Foundations of Digital Games*, pages 235–237.

McCoy, J., Treanor, M., Samuel, B., Tearse, B., Mateas, M., and Wardrip-Fruin, N. (2010a). Authoring game-based interactive narrative using social games and comme il faut. In *Proceedings of the 4th International Conference & Festival of the Electronic Literature Organization: Archive & Innovate*, volume 50. Citeseer.

McCoy, J., Treanor, M., Samuel, B., Tearse, B., Mateas, M., and Wardrip-Fruin, N. (2010b). Comme il faut 2: A fully realized model for socially-oriented gameplay. In *Proceedings of the Intelligent Narrative Technologies III Workshop*, pages 1–8.

McCoy, J., Treanor, M., Samuel, B., Wardrip-Fruin, N., and Mateas, M. (2011b). Comme il faut: A system for authoring playable social models. In *AIIDE*.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, 22(3):276–282.

Meehan, J. R. (1977). Tale-spin, an interactive program that writes stories. In *Ijcai*, volume 77, pages 91–98.

Miles, M. B., Huberman, A. M., Huberman, M. A., and Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook*. sage.

Miller, M. K. and Summers, A. (2007). Gender differences in video game characters' roles, appearances, and attire as portrayed in video game magazines. *Sex roles*, 57(9):733–742.

Millington, I. and Funge, J. (2018). *Artificial intelligence for games*. CRC Press.

Mirel, B. (1998). "applied constructivism" for user documentation: Alternatives to conventional task orientation. *Journal of business and technical communication*, 12(1):7–49.

Molleindustria (2014). To build a better mousetrap.

Morgan, D. L. (1993). Qualitative content analysis: a guide to paths not taken. *Qualitative health research*, 3(1):112–121.

Mori, G., Thue, D., and Schiffel, S. (2022). Em-glue: A platform for decoupling experience managers and environments. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 18, pages 266–274.

Morrison, H. and Martens, C. (2018). Making first impressions: A playable model of cross-cultural trust building. In *INTWICED@ AIIDE*.

Morrissette, J. J. (2017). Glory to arstotzka: Morality, rationality, and the iron cage of bureaucracy in papers, please. *Game Studies*, 17(1).

Moulin, B. and Brahim, C.-D. (1996). An overview of distributed. *Foundations of distributed artificial intelligence*, 9:1.

Nardone, G. (2014). The sims 4: Colors, objects, clothing & emotions.

Neo4j (2012). Neo4j - the world's leading graph database.

Nintendo, S. (2001). Animal crossing - stategywiki.

Nintendo EAD (2020). Animal crossing franchise.

Ortony, A., Clore, G. L., and Collins, A. (2022). *The cognitive structure of emotions*. Cambridge university press.

P. Carvalho, L., Jackson, D., Guerreiro, T., Guan, Y., and Montague, K. (2021). Participatory action research and open source hardware appropriation for large scale in-the-wild studies. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, pages 511–513.

Pew Research (2017). Political typology quiz - where do you fit in the political typology?

Porteous, J., Cavazza, M., and Charles, F. (2010). Narrative generation through characters' point of view. In *AAMAS*, pages 1297–1304. Citeseer.

Potter, W. J. and Levine-Donnerstein, D. (1999). Rethinking validity and reliability in content analysis. *Academia.org*.

Prendinger, H. and Ishizuka, M. (2001). Social role awareness in animated agents. In *Proceedings of the fifth international conference on Autonomous agents*, pages 270–277. ACM.

Pynadath, D. V. and Marsella, S. C. (2005). Psychsim: Modeling theory of mind with decision-theoretic agents. In *IJCAI*, volume 5, pages 1181–1186.

Repository, U. M. L. (2021). Uci machine learning repository: Data sets and benchmarks.

Richiardi, M. G., Leombruni, R., Saam, N. J., and Sonnessa, M. (2006). A common protocol for agent-based social simulation. *Journal of artificial societies and social simulation*, 9.

Riedl, M. O. (2016). Computational narrative intelligence: A human-centered goal for artificial intelligence. *arXiv preprint arXiv:1602.06484*.

Riedl, M. O. and Harrison, B. (2016). Using stories to teach human values to artificial agents. In *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence*.

Roberts, D. L. and Isbell, C. L. (2007). Desiderata for managers of interactive experiences: A survey of recent advances in drama management. In *Proceedings of the First Workshop on Agent-Based Systems for Human Learning and Entertainment (ABSHLE 07)*.

Rowe, J. P., Ha, E. Y., and Lester, J. C. (2008). Archetype-driven character dialogue generation for interactive narrative. In *International Workshop on Intelligent Virtual Agents*, pages 45–58. Springer.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145.

Ryan, J. (2016). Diegetically grounded evolution of gameworld languages. *Proc. Procedural Content Generation*.

Ryan, J. (2018). *Curating simulated storyworlds*. University of California, Santa Cruz.

Ryan, J., Brothers, T., Mateas, M., and Wardrip-Fruin, N. (2016a). Juke joint: characters who are moved by music. In *Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Ryan, J., Mateas, M., and Wardrip-Fruin, N. (2016b). A simple method for evolving large character social networks. *5th Workshop on Social Believability in Games at International Joint Conference of DiGRA and FDG*.

Ryan, J. O., Summerville, A., Mateas, M., and Wardrip-Fruin, N. (2015). Toward characters who observe, tell, misremember, and lie. In *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference*.

Samuel, B., Reed, A. A., Maddaloni, P., Mateas, M., and Wardrip-Fruin, N. (2015). The ensemble engine: Next-generation social physics. In *Proceedings of the Tenth International Conference on the Foundations of Digital Games (FDG 2015)*, pages 22–25.

Sengers, P., Boehner, K., David, S., and Kaye, J. (2005). Reflective design. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility*, pages 49–58.

Shapiro, D., Tanenbaum, K., McCoy, J., LeBron, L., Reynolds, C., Stern, A., Mateas, M., Ferguson, B., Diller, D., Moffitt, K., et al. (2015). Composing social interactions via social games. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 573–580.

Si, M., Marsella, S. C., and Pynadath, D. V. (2005). Thespian: Using multi-agent fitting to craft interactive drama. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '05, page 21–28, New York, NY, USA. Association for Computing Machinery.

Siebers, P.-O., Macal, C. M., Garnett, J., Buxton, D., and Pidd, M. (2010). Discrete-event simulation is dead, long live agent-based simulation! *Journal of Simulation*, 4(3):204–210.

Speiginer, G., MacIntyre, B., Bolter, J., Rouzati, H., Lambeth, A., Levy, L., Baird, L., Gandy, M., Sanders, M., Davidson, B., et al. (2015). The evolution of the argon web framework through its use creating cultural heritage and community–based augmented reality applications. In *International Conference on Human-Computer Interaction*, pages 112–124. Springer.

Spinuzzi, C. (2005). The methodology of participatory design. *Technical communication*, 52(2):163–174.

Stangor, C. (2020). Defining social psychology: History and principles. *Principles of Social Psychology-1st International Edition*.

Stieg, C. (2020). Why video games like the sims and animal crossing are so soothing right now.

Striner, A., Azad, S., and Martens, C. (2019). A spectrum of audience interactivity for entertainment domains. In *International Conference on Interactive Digital Storytelling*, pages 214–232. Springer.

Studios, L. (2018). Rimworld.

Summerville, A. and Samuel, B. (2020). Kismet: a small social simulation language. In *Summerville, A., Samuel, B.(2020, September). Kismet: a Small Social Simulation Language. In: 2020 International Conference on Computational Creativity (ICCC).(Casual Creator Workshop). ACC.*

Swartout, W., Hill, R., Gratch, J., Johnson, W. L., Kyriakakis, C., LaBore, C., Lindheim, R., Marsella, S., Miraglia, D., and Moore, B. (2006). Toward the holodeck: Integrating graphics, sound, character and story. Technical report, University Of Southern California Marina Del Rey CA Inst For Creative Technologies.

Tao, F., Zhang, H., Liu, A., and Nee, A. Y. (2018). Digital twin in industry: State-of-the-art. *IEEE Transactions on industrial informatics*, 15(4):2405–2415.

Thomas, F., Johnston, O., and Rawls, W. (1981). *Disney animation: The illusion of life*, volume 4. Abbeville Press New York.

Thue, D. and Bulitko, V. (2018). Toward a unified understanding of experience management. In *Fourteenth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Togelius, J., Yannakakis, G. N., Karakovskiy, S., and Shaker, N. (2013). Assessing believability. In *Believable bots*, pages 215–230. Springer.

Tosic, P. T. and Agha, G. A. (2004). Towards a hierarchical taxonomy of autonomous agents. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, volume 4, pages 3421–3426. IEEE.

Treanor, M., McCoy, J., and Sullivan, A. (2015). Social play in non-player character dialog. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 11.

Von Csefalvay, C. (2023). *Computational Modeling of Infectious Disease: With Applications in Python*. Elsevier.

Wang, F.-Y., Carley, K. M., Zeng, D., and Mao, W. (2007). Social computing: From social informatics to social intelligence. *IEEE Intelligent systems*, 22(2):79–83.

Wang, S.-W., Huang, C.-Y., and Sun, C.-T. (2014). Modeling self-perception agents in an opinion dynamics propagation society. *Simulation*, 90(3):238–248.

Ware, S. and Eger, M. (2022). Closing remarks AIIDE 2022. Available at: https://youtu.be/FmThZi7T5E0?t=117.

Ware, S. G., Garcia, E. T., Shirvani, A., and Farrell, R. (2019). Multi-agent narrative experience management as story graph pruning. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 15, pages 87–93.

Warpefelt, H. (2016). *The Non-Player Character: Exploring the believability of NPC presentation and behavior*. PhD thesis, Department of Computer and Systems Sciences, Stockholm University.

Weiss, R. S. (1991). The attachment bond in childhood and adulthood. *Attachment across the life cycle*, 8:66–76.

Wessells, A. T. (2007). Reassembling the social: An introduction to actor-network-theory by bruno latour: A review of:“reassembling the social: An introduction to actor-network-theory. bruno latour. new york: Oxford university press, 2005. 312 pages. isbn: 0199256047”.

with Code, P. (2021). Papers with code - image classification.

Wooldridge, M. and Jennings, N. R. (1994). Agent theories, architectures, and languages: a survey. In *International Workshop on Agent Theories, Architectures, and Languages*, pages 1–39. Springer.

Yannakakis, G. N. (2012). Game ai revisited. In *Proceedings of the 9th conference on Computing Frontiers*, pages 285–292. ACM.

Yannakakis, G. N. and Paiva, A. (2014). Emotion in games. *Handbook on affective computing*, 2014:459–471.

Young, R. M. (2000). Creating interactive narrative structures: The potential for ai approaches. *Psychology*, 13:1–26.

Young, R. M., Ware, S. G., Cassell, B. A., and Robertson, J. (2013). 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.

Zook, A., Lee-Urban, S., Riedl, M. O., Holden, H. K., Sottilare, R. A., and Brawner, K. W. (2012). Automated scenario generation: toward tailored and optimized military training in virtual environments. In *Proceedings of the international conference on the foundations of digital games*, pages 164–171.

Zoric, G., Smid, K., and Pandzic, I. S. (2007). Facial gestures: taxonomy and application of non-verbal, non-emotional facial displays for embodied conversational agents. *Conversational Informatics: An Engineering Approach*, pages 161–182.

APPENDICES